Salivary gland tumors: profiling and biomarker discovery

Matse, J.H.

2017

document version
Publisher's PDF, also known as Version of record

Salivary gland tumors:
Profiling and Biomarker Discovery

Johannes Hendrikus Matse

The studies described in this thesis were carried out at the departments of Oral and Maxillofacial Surgery/ Oral Pathology at the VU University Medical Center and the department of Oral Biochemistry at Academic Centre for Dentistry , Amsterdam, The Netherlands.

ISBM:
Printed & Lay Out by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, ‘s-Hertogenbosch
Cover designed by Alan E. Bailey (Bridge City, TX, USA)

VRIJE UNIVERSITEIT

Salivary gland tumors: profiling and biomarker discovery

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Tandheelkunde
op dinsdag 23 mei 2017 om 11.45 uur
in het aula van de universiteit,
De Boelelaan 1105

Door

Johannes Hendrikus Matse

geboren te Utrecht

promotoren:	prof.dr. E. Bloemena
	prof.dr. E.C.I. Veerman
copromotor:	dr. J.G.M. Bolscher

“No mercy” – Bertran R. Matse

Contents

1.

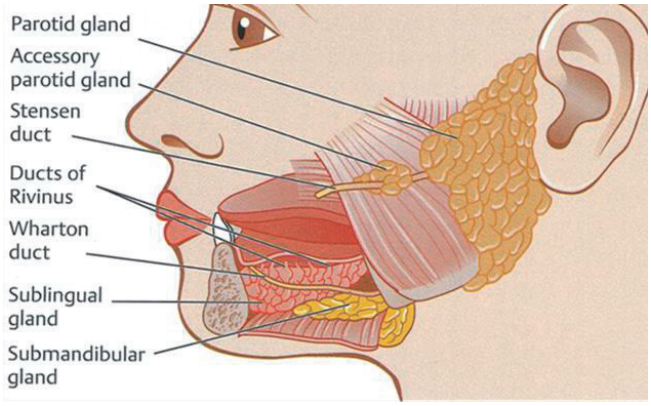
General Introduction

General Introduction

Saliva is composed of a mixture of oral fluids from the major salivary (submandibular 65%, parotid 23% and sublingual 4%) and minor salivary glands (8%)[1]. Saliva serves digestive functions through the delivery of lipase and amylase, lubricates the oral cavity with a mucin-rich film, maintains a neutral pH through its buffering capacity, re-mineralizes the enamel of the teeth, cleans the oral cavity, stimulates wound-healing and has antimicrobial properties[1]. Furthermore, saliva is vital to the full enjoyment of the pleasures of life, such as kissing, talking and tasting foods[2]. Saliva also contains a small amount of serum. Small neutral molecules from the serum enter the oral cavity by passive diffusion from a dense bed of capillaries that surround and bathe the salivary glands. Furthermore, serum components leak into the mouth through damaged or inflamed mucosa. This provides saliva with a low concentration of many molecules which are also present in the systemic circulation, making saliva a microcosmic mirror of the body[3].

The production and secretion of saliva is facilitated by the salivary glands. These exocrine organs are comprised of the three paired major glands: the parotid, submandibular and sublingual glands (Figure 1), and the minor salivary glands. The latter glands are numerous and are widely distributed throughout the mucosa of the upper aerodigestive tract[1].

Figure 1: Position of the major salivary gland in the human head. The three main pairs of salivary glands are the parotid glands, the sublingual glands and the submandibular glands. (Picture adopted from Bradley and Guntinas-Lichius[1]).



The functional structures of the salivary glands are the acini and the related intercalated and striated ducts (Figure 2). The acini are the secretory units of the salivary gland which secrete a fluid comprised of water, electrolytes, (glyco)proteins and enzymes and are categorized as serous, mucous or seromucous (mixed) acini[1]. The parotid gland consists almost

exclusively of serous acini; while the sublingual gland consists primarily of mucous acini. The submandibular gland, in which 90% of the acini are of the serous type, is classified as a mixed gland. The acini and the intercalated ducts are surrounded by myoepithelial cells or basket cells. The myoepithelial cells around the acini have a star-like, dendritic appearance and form a basket around the acini, while the myoepithelial cells around the intercalated duct have an elongated form. Myoepithelial cells have the ability to contract, squeezing the acini to aid in saliva secretion and to speed its outflow. A series of ducts, the smallest of which is the intralobular intercalated duct, transport the saliva. The secretion of saliva is regulated by the autonomic nervous system. Activity of the parasympathetic nervous system causes a rich flow of saliva, while activity of the sympathetic nervous system causes only a small-volume of a protein-rich secretion[4]. The secretion of saliva is under control of both the parasympathetic and sympathetic branch of the autonomic nervous system[4]. Because of the large-volume response resulting from activity of the parasympathetic nervous system, protein concentration is lower than in saliva stimulated by the sympathetic nervous system. The secretory activity of salivary glands may also be regulated through gastrointestinal hormones. Serous acinic cells and duct cells of the major salivary glands develop receptors for the gastrointestinal hormones gastrin, cholecystokinin, and melatonin[1]. These hormones activate protein secretion without any accompanying fluid secretion as shown in parotid glands of rats as well as in human gland tissue *in vitro*[5].

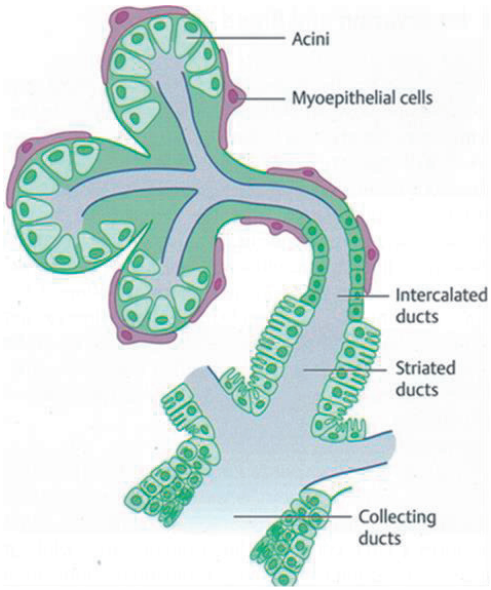


Figure 2: Schematic diagram of the structures of the salivary gland, showing the arrangements of the acini and the layout of the ducts (Picture adopted from Bradley and Guntinas-Lichius[1]).

Salivary gland tumors

The World Health Organization classification of 2005 recognizes 37 subtypes (13 benign and 24 malignant) salivary gland tumors. These tumors exhibit a wide variety of histopathological appearances within or between tumors[6]. This presents difficulties for classifying and diagnosing salivary gland tumors. Salivary gland tumors represent 2-6.5% of all head and neck neoplasms, and their global annual incidence is 0.4-13.5/100,000 of cases[6]. The parotid gland is the most commonly affected gland. Eighty percent of all salivary gland tumors occur in this gland, whereas only 7-11% occurs in the submandibular gland and less than 1% occurs in the sublingual gland[6]. The proportion of malignant tumors, however, varies greatly by site. Malignant tumors comprise 15-32% of parotid tumors, 41-45% of submandibular tumors, and 70-90% of sublingual tumors. The average age of patients with either benign or malignant salivary gland tumors is about 46 years. The most common salivary gland tumor is the pleomorphic adenoma. This benign subtype accounts for almost 60% of all benign tumors in the major salivary glands. Of the most common malignant tumor subtypes, mucoepidermoid carcinoma and the adenoid cystic carcinoma accounts for approximately 50% and 10% of all malignant tumors in the major salivary glands, respectively[6].

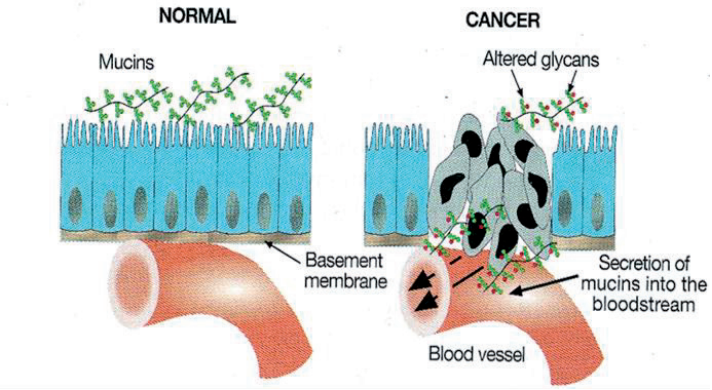
The etiology of salivary gland tumors is so far unknown. Putative risk factors include: cigarette smoking, genetic predisposition, viral infections, rubber manufacturing, plumbing, some types of woodworking, asbestos mining as well as exposure to nickel compounds[7]. The only well-established risk factor is ionizing radiation. Atomic bomb survivors and cancer patients treated by radiation show a substantially higher risk of developing salivary gland tumors[8]. There is a strong association between Warthin tumor and cigarette smoking, with Warthin tumor occurring 8 times more often in smokers than in non-smokers. Irritants in tobacco smoke may cause metaplasia in the parotid gland. The association with tobacco use may explain the higher incidence of Warthin tumor in males[6].

Mucins and mucin expression in cancer

Mucins are glycoproteins with a high molecular weight. They contain long carbohydrate sidechains attached to serine or threonine residues. Their large size in combination with the abundant amount of carbohydrates are responsible for the gel-forming and visco-elastic properties of mucins[9]. Mucins are generally classified as membrane-bound mucins (MUC1, MUC3, MUC4, MUC12, MUC13, MUC15-MUC17, MUC20, and MUC21) or as secreted mucins (MUC2, MUC5AC, MUC5B, MUC6-MUC8 and MUC19)[9].

Mucins are secreted from or located on the apical borders of normal epithelial cells. In malignant cells, however, there is a loss of polarity that occurs in association with the activity of a proliferation and survival program. With the loss of polarity, membrane-bound mucins are transiently repositioned over the entire cell membrane (Figure 3)[10].

Figure 3: Loss of normal topology and polarization of epithelial cells in cancer results in abnormal expression of mucins (Picture adopted from Varki and Cummings et al[9]).



Overexpression and repositioning of MUC1 and MUC4 have been reported for different type of cancers[11-20]. Previous studies describing the overexpression of MUC1 in human breast carcinomas led to the finding that the extracellular part of MUC1 (MUC1-N subunit) is detectable at increased levels in the serum of patients with breast cancer[21,22]. Expression of the MUC1-C subunit in lung and breast cancer patients is associated with a significant decrease in disease-free and overall survival rates[11]. Overexpression of MUC4 has been found in carcinomas of the pancreas[12], gall bladder[23], ovary[13], breast[14], and lung[15]. MUC4 expression is, however, reduced in prostate carcinomas[15]. In lung cancers, overexpression of MUC4 has been used to distinguish lung adenocarcinomas from epithelial mesothelioma[17]. Therefore, MUC4 expression has been demonstrated to be a good marker for prognosis in human tumors, depending on the tumor type.

Mucin expression in salivary gland tumors has been studied mostly in mucoepidermoid carcinomas, followed by pleomorphic adenomas and adenoid cystic carcinomas. The expression of the membrane-associated mucin MUC1 has been investigated in different subtypes of salivary gland tumor. Mucoepidermoid carcinoma (as well as acinic cell carcinoma, adenoid cystic carcinoma, salivary duct carcinoma and pleomorphic adenoma) were all positive for MUC1. Besides an increased expression of MUC1, the localization of MUC1 changed in these tumors from being mainly expressed on the apical membrane to being expressed over the entire cell membrane. The overexpression of MUC1 is associated with high recurrence rates in mucoepidermoid carcinoma and in pleomorphic adenoma[18-20,24].

The abnormal MUC4 expression is associated with tumor differentiation in mucoepidermoid carcinoma. A high MUC4 expression in mucoepidermoid carcinoma is mainly related to low-grade tumors, lower recurrence rates, a longer disease-free interval and a reduced risk of death from the disease[20,24, 25].

Expression of MUC5B, a secretory mucin that is normally expressed by (sero)mucous salivary glands, has been investigated in 3 salivary gland subtypes: cystadenoma, mucoepidermoid carcinoma and salivary duct carcinoma. In one case study, it was reported that cystadenoma (of the palate) was negative for MUC5B[26]. On the other hand, 4/5 salivary duct carcinoma s[27] and 33/40 mucoepidermoid carcinoma samples were MUC5B positive. Low grade mucoepidermoid carcinomas were positive for MUC5B, whereas most high grade were negative[20, 24].

MUC5AC, which is not expressed by normal salivary glands, was expressed in approximately 65% of the mucoepidermoid carcinomas, mostly of the low grade type. Most high grade mucoepidermoid carcinomas did not express MUC5AC[20,24]. An overview of immuno-reactivity of mucins in salivary gland tumors can be found in Table 1.

Table 1: Immuno-reactivity of mucins in salivary gland tumors.

Salivary gland tumor subtype	MUC1 (%)	MUC4 (%)	MUC5AC (%)	MUC5B (%)
Mucoepidermoid carcinoma	96/114 (84)	109/131 (83)	92/123(75)	33/40 (83)
Acinic cell carcinoma	19/20 (90)	0/8 (0)	0/11 (0)	NT
Adenoid cystic carcinoma	60/60 (100)	1/2 (50)	0/20 (0)	NT
Salivary duct carcinoma	2/2 (100)	1/1 (100)	2/6 (33)	4/5 (80)
Pleomorphic adenoma	106/158 (67)	7/51 (14)	6/49 (12)	NT
Warthins Tumor	28/29 (97)	3/3 (100)	NT	NT

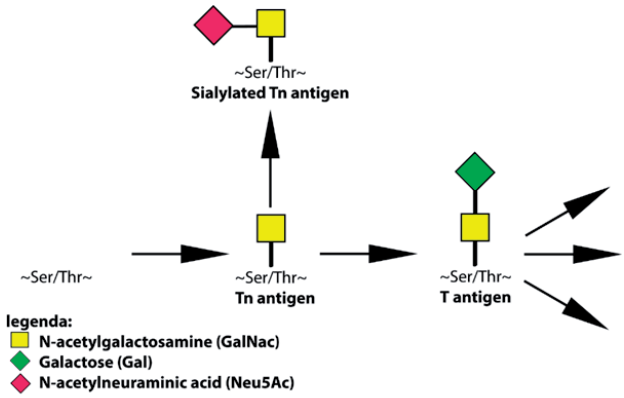
NT. Not Tested. (Table adopted from F. Mahomed [28])

Aberrant expression of O-linked carbohydrate antigens in cancer

Mucins can carry hundreds of highly diverse carbohydrate sidechains attached to their polypeptide backbone. The synthesis of these carbohydrate sidechains occurs co- and post-translationally, and carbohydrates are mainly synthesized by the sequential action of a variety of glycosyltransferases[9]. It starts with the addition of an N-acetylgalactosamine (GalNAc) residue to a serine or threonine of the polypeptide backbone, forming the Tn antigen, followed by a stepwise addition of successive carbohydrate residues (Figure 4). This will give rise to a structurally highly diverse set of carbohydrate sidechains which may contain > 20 carbohydrate residues.

When the glycosylation machinery is disrupted, e.g. in cancer cells, the stepwise elongation of the oligosaccharides can be prematurely terminated. As a result, mucins expressed by cancer cells often contain an anomalously high number of short immature carbohydrate antigens, such as the Tn antigen (GalNAcα1-O-Ser/Thr), sialyl-Tn (sialylα2-6GalNAcα1-O-Ser/Thr), and T antigen (Galβ1-3GalNAcα1-O-Ser/Thr) (Figure 4).

Figure 4: Initial steps of O-linked glycosylation. In cancer, incomplete glycosylation in the O-linked pathway results in expression of the Tn antigen, the sialylated Tn antigen ("dead-end" structure), or the T antigen. Ser: serine, Thr: threonine.



The expression of such simple mucin-type carbohydrate antigens has been observed in several cancers such as breast and colon cancer[29,30]. Expression of sialyl-Tn is in breast cancer associated with an unfavorable prognosis[29]. Increased Tn and sialyl-Tn expression in colon and breast cancers is associated with advanced cancer, highly proliferative and invasion tumors, metastasis and a poor clinical outcome[29]. In salivary gland tumors, aberrant expression of Tn and sialyl-Tn antigens was observed in the cytoplasm of glandular cells, on luminal membranes and on mucinous content in almost all mucoepidermoid carcinomas and adenocarcinomas[31]. Patients with Tn-positive malignant salivary gland tumors have an increased risk of loco-regional recurrence and early deaths[31]. So far, no other associations have been found between the expression of simple mucin-type carbohydrate antigens and other subject factors such as gender, tumor size, prognosis or survival.

Tumors and micro-array comparative genomic hybridization

The biological behavior of cancer is thought to be associated primarily with genetic aberrations in the tumor cells themselves. Changes in DNA copy number (gains/losses), as well as structural alterations of chromosomes are implicated in the development and/or progression of different malignancies[32-35].

Micro-array comparative genomic hybridization (arrayCGH) is a technique which allows for efficiently scanning of the entire genome for variations in DNA copy number aberrations. The technique uses an array of small DNA segments as targets for analysis (Figure 5). These may vary in size: from oligonucleotides manufactured to represent areas of interest (25–85 base pairs) to genomic clones such as bacterial artificial chromosomes (80,000–200,000 base pairs).

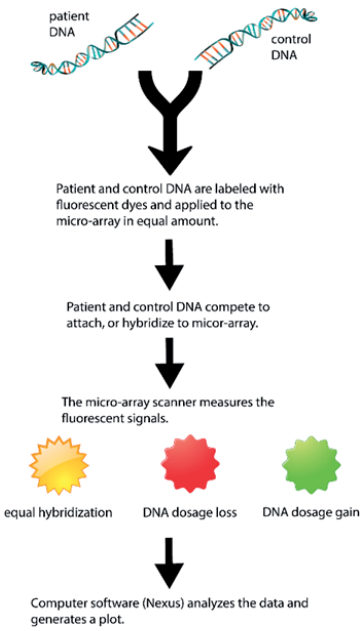


Figure 5: Steps in micro-array technique. First, control and tumor DNA are isolated and labeled with the fluorescent labels, Cy5 and Cy3 The Cy3-labeled tumor and the Cy5-labeled control DNA are then hybridized together on a micro-array slide. The relative hybridization intensity of the reference and test signals at a certain location is then proportional to the relative copy number of those sequences in the test and reference genomes. If the reference genome is normal, then increases and decreases in the intensity ratio directly indicate DNA copy number variation in the genome of the test population. Yellow means no genomic aberrations, red means that there is a DNA deletion and green means that there is a DNA amplification.

Associations between copy number aberration and prognosis have been found for a variety of tumors, including: prostate cancer[38], breast cancer[39] and gastric cancer[40]. Little is known about recurring copy number aberration in mucoepidermoid carcinoma. One study that looked at recurring copy number aberration found that high grade mucoepidermoid carcinomas had significantly more and larger copy number aberrations compared to low grade mucoepidermoid carcinomas[41]. Frequently gained regions found were: 3q26.1-q28, 3q33.3-q34.3, 5pter-p15.31, 8q24.3 and19p13.2-p13.11 (in which p represents the short arm of a chromosome and q represents the longs arm). Most found lost regions were: 4p, 5q13.2-q15, 6q22.1-q23.1, 8pter-p12.1, 9p21.3 and 18q12.2-qter[41]. These copy number aberrations may result in chromosomal translocations in which a chromosomal segment is moved from one position to another, either within the same chromosome or to another chromosome. These translocated genes encode novel proteins which can alter the expression of proteins. The (abnormal) expression of these proteins can positively influence signaling pathways involved in tumor progression[42].

Several recurring translocations have been identified in salivary gland tumors[1]. A subgroup of mucoepidermoid carcinomas is characterized by a t(11;19)(q21;p13) translocation[43-48]. The translocation results in a fusion of parts of the *MAML2* (11q21) and *METC1* (19p13) genes. The t(11;19)(q21;p13) translocation is shared by acute leukemia, and an apparently identical rearrangement has been identified in Warthin tumor[49]. Apart from Warthin tumor and mucoepidermoid carcinomas, it has not been demonstrated in other salivary gland tumors[50].

First described by Nordkvist et al[51], it was Tonon et al[43] who characterized the *METC1* (mucoepidermoid carcinoma translocated-1) (also known as *CRTC1*, *TORC1* and *WAMTP1*) and *MAML2* (master-mind-like 2) as the underlying pathogenetic event in the majority of mucoepidermoid carcinomas. The *METC1-MAML2* fusion protein can activate both cAMP-CREB target and Notch signaling targets[52]. This will disrupt both cell cycle and differentiation functions. The translocation is detected in approx. 55% of the mucoepidermoid carcinomas, with a higher prevalence in low or intermediate grade mucoepidermoid carcinomas (75%) compared to high grade mucoepidermoid carcinomas (46%)[43-48]. Fusion-positive cases show significantly better survival than fusion-negative cases[53, 54].

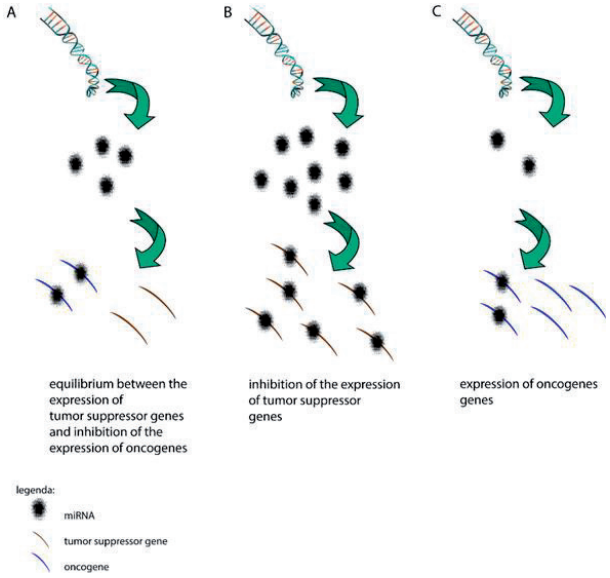
Jee et al[41] found that fusion-negative cases had significantly more copy number aberrations compared to fusion-positive tumors; the fusion-negative cases were mostly high grade mucoepidermoid carcinomas and most of the fusion-positive cases were low grade mucoepidermoid carcinomas. Anzick et al[55] found that fusion-negative cases with poor prognosis often showed a *CDKN2A* deletion or methylation. The deletion or (hyper) methylation of *CDKN2A* has been demonstrated to be an early tumorigenic event in other types of carcinomas including squamous cell carcinomas of the head and neck and lung cancer[56-58]. However, little is known about the role of *CDKN2A* in mucoepidermoid carcinoma or in other salivary gland tumors.

Micro-RNA

Micro-RNAs (miRNAs) are small non-coding RNAs consisting of 19-25 nucleotides. A miRNA can have dozens of target genes which are involved in various cellular processes such as cell differentiation, proliferation and survival[59]. MiRNAs are expressed in 12 bodily fluids among which are: urine, saliva, tears and seminal fluid[60]. The expression of certain miRNAs is tissue and cell specific. Mature miRNA are secreted into bodily fluids via small vesicles called exosomes. The majority of the miRNAs found in saliva are in these exosomes.

Altered levels of miRNA expression have been implicated in the etiology of cancer [61,62]. An increase of onco-miRNA expression can result in the translation inhibition of tumor-suppressor genes (Figure 6B). On the other hand, the decreased expression of anti-onco-miRNA can result in the increased expression of oncogenes (Figure 6C). MiRNAs can do this by binding to the complementary sequences in the 3' UTR of multiple target messenger RNAs of a tumor-suppressor or oncogenes. When a miRNA fits perfectly, the mRNA is degraded. If the miRNA fits imperfectly, the translation of the mRNA is repressed[61,62].

Figure 6: The role miRNA play in the etiology of cancer. A. "normal" situation in which there is an equilibrium between the expression of tumor suppressor genes and the inhibition of oncogenes. B-C miRNA expression in cancer. Due to chromosomal changes more or less miRNA are being expressed, which can lead to inhibition of tumor suppressor genes and an expression of oncogenes.



Because of the big role they play in the etiology of cancer, miRNA expression in serum and saliva have become an interesting research topic in cancer diagnostics. As in most diseases, early detection and diagnosis can lead to a higher survival rates. Therefore the ability to evaluate physiological conditions, follow the progression of the disease and monitor post-treatment therapeutic results through a noninvasive method is one of the primary objectives in the field of healthcare research. Saliva can potentially be used as a specimen for diagnosis because of its exchange with substances existing in human serum[63]. There has been an increasing interest in the use of saliva as a non-invasive diagnostic medium, especially with the rise of technology that is increasingly able to measure the small concentrations of molecules present in saliva.

Several studies have investigated the utility of salivary miRNA as a tool for tumor diagnostics. In saliva samples from patients with esophageal cancer, the expression level of 3 miRNAs was significantly higher in whole saliva compared to whole saliva samples of the control group. Changes in miRNA profiles have also been described in saliva samples from patients with oral squamous cell carcinoma[64-67]. The expression level of miR-125a and miR-200a was significantly lower in saliva from patients than in healthy control subjects[66], while miR-31 was overexpressed in saliva from patients with oral squamous cell carcinoma [67].

SCOPE OF THE THESIS

This thesis describes the attempt to find molecular, non-invasive biomarkers for the profiling and diagnosis of salivary gland tumors. Because salivary gland tumors are a heterogenic group of tumors, diagnosing them is still very challenging. Therefore, using novel biomarkers as an additional tool can add to the accuracy of diagnoses of salivary gland tumors. Biomarkers can be found on DNA, RNA and protein levels, and all three levels were addressed: (1) on DNA level by cataloging genomic aberrations using arrayCGH, (2) on protein level by investigating the expression of mucins and mucin-associated carbohydrates by immunohistochemistry. (3) on RNA level by determining the miRNA expression levels in whole and parotid saliva by RT-qPCR and answers were sought to the following questions:

Chapter 2: Are mucins (MUC1, MUC4, MUC5AC, MUC5B) and mucin–type carbohydrate epitopes (Tn antigen, sialyl–Tn antigen, T antigen, Lewis^a and sulfo–Lewis^a) aberrantly expressed in mucoepidermoid carcinoma, and how does the expression relate tumor characteristics?

The expression patterns of mucin and mucin-type carbohydrate antigen expression patterns were investigated in mucoepidermoid carcinoma (n=45) and normal salivary gland (n=8) tissue using immunohistochemical analysis. Expression profiles were examined on where mucins and mucin-type carbohydrates were expressed. Additionally, we correlated tumor characteristics such as tumor grade.

Chapter 3: What are the recurrent copy number aberrations in mucoepidermoid carcinoma, and how do they correlate to histological grade and translocation status of the tumor?

A genome-wide copy number aberration analysis was conducted in mucoepidermoid carcinoma (n=27) using micro-array comparative-genomic-hybridization. Additionally, we determined the t(11;19)(q21;p13) translocation status of each tumor sample. We correlated recurrent copy number aberrations to the histological grade or the translocation status of the tumor to discover recurrent genomic regions that can classify these heterogenic tumors.

Chapter 4: Are there differences in the expression of miRNAs in whole saliva from patients with a parotid gland tumor and healthy controls, and can differences in miRNA expression levels be used as a predictive indicator? Are the discovered miRNA reflected in the parotid saliva from the patients' affected and non–affected parotid glands?

MiRNA expression in whole saliva from patients with a parotid gland tumor and healthy controls were profiled using RT-qPCR, and examined which of the miRNAs are differently expressed. The discovered miRNAs were validated in an independent sample set, and the diagnostic power of the discovered miRNAs was investigated using multivariate logistic

regression analysis. Additionally, the expression of the discovered miRNAs was researched in the parotid saliva from the patients' affected and non-affected parotid gland.

Chapter 5: Are there differences in the expression of miRNAs in whole saliva from patients with a benign or malignant parotid gland tumor, and can the expression levels be used as a predictive indicator for benign or malignant parotid gland tumor?
Differences in the expression of miRNA in whole saliva from patients with a benign parotid gland tumor and patients with a malignant parotid gland tumor were profiled using RT-qPCR. The expression of the discovered miRNAs was validated in an independent whole saliva sample set. Furthermore, the predictive power of the discovered miRNAs was investigated using multivariate logistic regression analysis.

Chapter 6: Can elastic net algorithms find group–based miRNAs that can distinguish whole saliva from patient with a benign or malignant parotid gland tumor? What are the targets for the discovered group–based miRNAs?
Using machine learning, the data from the discovery phase from chapter 4 was re-analyzed to find group-based miRNA that can predict the presence of a benign or a malignant parotid gland tumor. Additionally, a hypothetical model was created using literature to find validated targets that may be influenced by the expression of the discovered group-based miRNAs.

References

1. Bradley PJ, Guntinas-Lichius O. 2011. Salivary gland disorders and diseases: diagnosis and management. ER Carlson and RA Ord (Ed.) Thieme, Stuttgart/New York.
2. Tenovuo J. Antimicrobial function of human saliva--how important is it for oral health? Acta Odontologica Scandinavica. 1998; 56:250-6.
3. Miller CS, Foley JD, Bailey AL, et al. Current developments in salivary diagnostics. Biomarkers in Medicine. 2010; 4:171-89.
4. Ekström J. Autonomic control of salivary secretion. Proceedings of the Finnish Dental Society. 1989;85:323-31.
5. Aras HC, Ekström J. Melatonin-evoked in vivo secretion of protein and amylase from the parotid gland of the anaesthetised rat. Journal of Pineal Research. 2008; 45:413-21.
6. Barnes L, Eveson JW, Reichart P, et al. Pathology and Genetics of Head and Neck Tumours (IARC WHO Classification of Tumors) (IARC Press, Lyon, 2005).
7. Söderqvist F, Carlberg M, Hardell L. Use of wireless phones and the risk of salivary gland tumours: a case-control study. European Journal of Cancer Prevention. 2012; 21:576-9.
8. Takeichi N, Hirose F, Yamamoto H. Salivary gland tumors in atomic bomb survivors, Hiroshima, Japan. I. Epidemiologic observations. Cancer. 1976; 38:2462-8.
9. Varki A, Cummings RD, Esko JD, et al. Essentials of Glycobiology Ch. 11 and 44 (Cold Spring Harbour Laboratory Press, Cold Springs Harbour/New York, 2009).
10. Kufe D. Mucins in cancer: function, prognosis and therapy. Nature Review Cancer. 2009: 9:874-85
11. Khodarev NN, Pitroda SP, Beckett MA, et al. MUC1-induced transcriptional programs associated with tumorigenesis predict outcome in breast and lung cancer. Cancer Research. 2009; 69:2833-7.
12. Saitou M, Goto M, Horinouchi M, et al. MUC4 expression is a novel prognostic factor in patients with invasive ductal carcinoma of the pancreas. Journal of Clinical Pathology. 2005; 58:845-52.
13. Giuntoli RL 2nd, Rodriguez GC, Whitaker RS. Mucin gene expression in ovarian cancers. Cancer Research. 1998; 58:5546-50.
14. Rakha EA, Boyce RW, Abd El-Rehim D, et al. Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. Modern Pathology. 2005;18:1295-304.
15. Kwon KY, Ro JY, Singhal N, et al. MUC4 expression in non-small cell lung carcinomas: relationship to tumor histology and patient survival. Archives of Pathology & Laboratory Medicine. 2007; 131:593-8.
16. Singh AP, Chauhan SC, Bafna S, et al. Aberrant expression of transmembrane mucins, MUC1 and MUC4, in human prostate carcinomas. Prostate. 2006; 66:421-9.
17. Llinares K, Escande F, Aubert S, et al. Diagnostic value of MUC4 immunostaining in distinguishing epithelial mesothelioma and lung adenocarcinoma. Modern Pathology. 2004; 17:150-7.
18. Hamada T, Matsukita S, Goto M, et al. Mucin expression in pleomorphic adenoma of salivary gland: a potential role for MUC1 as a marker to predict recurrence. Journal of Clinical Pathology. 2004; 57:813-21.

19. Soares AB, Demasi AP, Altemani A, et al. Increased mucin 1 expression in recurrence and malignant transformation of salivary gland pleomorphic adenoma. Histopathology. 2011; 58:377-82.
20. Handra-Luca A, Lamas G, Bertrand JC, et al. MUC1, MUC2, MUC4, and MUC5AC expression in salivary gland mucoepidermoid carcinoma: diagnostic and prognostic implications. American Journal of Surgical Pathology. 2005; 29:881-9.
21. Kufe D, Inghirami G, Abe M, et al. Differential reactivity of a novel monoclonal antibody (DF3) with human malignant versus benign breast tumors. Hybridoma. 1984; 3:223-32.
22. Hayes DF, Sekine H, Ohno T, et al. Use of a murine monoclonal antibody for detection of circulating plasma DF3 antigen levels in breast cancer patients. Journal of Clinical Investigation. 1985; 75:1671-8.
23. Miyahara N, Shoda J, Ishige K, et al. MUC4 interacts with ErbB2 in human gallbladder carcinoma: potential pathobiological implications. European Journal of Cancer. 2008; 44:1048-56.
24. Alos L, Lujan B, Castillo M, et al. Expression of membrane-bound mucins (MUC1 and MUC4) and secreted mucins (MUC2, MUC5AC, MUC5B, MUC6 and MUC7) in mucoepidermoid carcinomas of salivary glands. American Journal of Surgical Pathology. 2005; 29:806-13.
25. Weed DT, Gomez-Fernandez C, Pacheco J, et al. MUC4 and ERBB2 expression in major and minor salivary gland mucoepidermoid carcinoma. Head Neck. 2004; 26:353-64.
26. Kusafuka K, Maeda M, Honda M, et al. Mucin-rich salivary duct carcinoma with signet-ring cell feature ex pleomorphic adenoma of the submandibular gland: a case report of an unusual histology with immunohistochemical analysis and review of the literature. Medical Molecular Morphology. 2012; 45:45-52.
27. Simpson RH, Prasad AR, Lewis JE, et al. Mucin-rich variant of salivary duct carcinoma: a clinicopathologic and immunohistochemical study of four cases. American Journal of Surgical Pathology. 2003; 27:1070-9.
28. Mahomed F. Recent advances in mucin immunohistochemistry in salivary gland tumors and head and neck squamous cell carcinoma. Oral Oncology. 2011; 47:797-803
29. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. EMBO Reports. 2006; 7:599-604
30. Nakagoe T, Sawai T, Tuji T, et al. Prognostic value of expression of sialosyl-Tn antigen in colorectal carcinoma and transitional mucosa. Digestive Diseases and Science. 2002; 47:322-30.
31. Therkildsen MH, Anderson LJ, Christensen M, et al. Salivary gland carcinomas: prognostic significance of simple mucin-type carbohydrate antigens. Oral Oncology. 1998; 34:44-51
32. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992; 258:818-21.
33. Mertens F, Johansson B, Höglund M, et al. Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms. Cancer Research. 1997; 57:2765-80.
34. Knuutila S, Björkqvist AM, Autio K, et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. American Journal of Pathology. 1998; 152:1107-23.

35. Smetana J, Fröhlich J, Vranová V, et al. Oligonucleotide-based Array CGH as a Diagnostic Tool in Multiple Myeloma Patients. Klinická onkologie. 2011; 24 (Suppl): S43–S48.
36. Knuutila S, Aalto Y, Autio K, et al. DNA copy number losses in human neoplasms. American Journal of Pathology. 1999; 155:683-94.
37. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nature Genetics. 1998; 20:207-11.
38. Paris PL, Andaya A, Fridlyand J, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Human Molecular Genetics. 2004; 13:1303-13.
39. Callagy G, Pharoah P, Chin SF, et al. Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. Journal of Pathology. 2005; 205:388-96.
40. Weiss MM, Kuipers EJ, Postma C, et al. Genomic alterations in primary gastric adenocarcinomas correlate with clinicopathological characteristics and survival. Cellular Oncology. 2004; 26:307-17.
41. Jee KJ, Persson M, Heikinheimo K, et al. Genomic profiles and CRTC1-MAML2 fusion distinguish different subtypes of mucoepidermoid carcinoma. Modern Pathology. 2013; 26:213-22.
42. Stenman G. Fusion oncogenes and tumor type specificity--insights from salivary gland tumors. Seminars in Cancer Biology. 2005; 15:224-35.
43. Tonon G, Modi S, Wu L, et al. t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. Nature Genetics. 2003; 33:208-13.
44. Behboudi A, Enlund F, Winnes M, et al. Molecular classification of mucoepidermoid carcinomas-prognostic significance of the MECT1-MAML2 fusion oncogene. Genes Chromosomes Cancer. 2006; 45:470-81.
45. Fehr A, Röser K, Heidorn K, et al. A new type of MAML2 fusion in mucoepidermoid carcinoma. Genes Chromosomes Cancer. 2008; 47:203-6.
46. Verdorfer I, Fehr A, Bullerdiek J, et al. Chromosomal imbalances, 11q21 rearrangement and MECT1-MAML2 fusion transcript in mucoepidermoid carcinomas of the salivary gland. Oncology Reports. 2009; 22:305-11.
47. Seethala RR, Dacic S, Cieply K, et al. A reappraisal of the MECT1/MAML2 translocation in salivary mucoepidermoid carcinomas. American Journal of Surgical Pathology. 2010; 34:1106-21.
48. Schwarz S, Stiegler C, Müller M, et al. Salivary gland mucoepidermoid carcinoma is a clinically, morphologically and genetically heterogeneous entity: a clinicopathological study of 40 cases with emphasis on grading, histological variants and presence of the t(11;19) translocation. Histopathology. 2011; 58:557-70.
49. Winnes M, Enlund F, Mark J, et al. The MECT1-MAML2 gene fusion and benign Warthin's tumor: is the MECT1-MAML2 gene fusion specific to mucuepidermoid carcinoma? Journal of Molecular Diagnostics. 2006; 8:394-5.
50. Chenevert J, Barnes LE, Chiosea SI. Mucoepidermoid carcinoma: a five-decade journey. Virchows Archives. 2011; 458:133-40.
51. Nordkvist A, Gustafsson H, Juberg-Ode M, et al. Recurrent rearrangements of 11q14-22 in mucoepidermoid carcinoma. Cancer Genetics and Cytogenetics. 1994; 74:77-83.

52. Wu L, Liu J, Gao P, et al. Transforming activity of MECT1-MAML2 fusion oncoprotein is mediated by constitutive CREB activation. The EMBO Journal. 2005; 24:2391-402.
53. Okabe M, Miyabe S, Nagatsuka H, et al. MECT1-MAML2 fusion transcript defines a favorable subset of mucoepidermoid carcinoma. Clinical Cancer Research. 2006; 12:3902-7.
54. Bell D, El-Naggar AK. Molecular heterogeneity in mucoepidermoid carcinoma: conceptual and practical implications. Head Neck Pathology. 2013; 7:23-7.
55. Anzick SL, Chen WD, Park Y, et al. Unfavorable prognosis of CRTC1-MAML2 positive mucoepidermoid tumors with CDKN2A deletions. Genes Chromosomes Cancer. 2010; 49:59-69.
56. Rocco JW, Sidransky D. p16(MTS-1/CDKN2/INK4a) in cancer progression. Experimental Cell Research. 2001; 264:42-55.
57. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nature Review Cancer. 2006; 6:107-16.
58. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. New England Journal of Medicine. 2008; 358:1118-28.
59. itchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences of the USA. 2008; 105:10513-8.
60. Weber JA, Baxter DH, Zhang S, et al. The microRNA spectrum in 12 body fluids. Clinical Chemistry. 2010; 56:1733-41.
61. Chen PS, Su JL, Hung MC. Dysregulation of MicroRNAs in cancer. Journal of Biomedical Science. 2012; 19:90.
62. Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. Journal of Clinical Oncology. 2009; 27:5848-56
63. Cuevas-Córdoba B, Santiago-García J. Saliva: A Fluid of Study for OMICS. OMICS. 2014; 18:87-97.
64. Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009; 15:5473-7.
65. Brinkmann O, Wong DT. Salivary transcriptome biomarkers in oral squamous cell cancer detection. Advances in Clinical Chemistry. 2011; 55:21-34.
66. Yoshizawa JM, Wong DT. Salivary microRNAs and oral cancer detection. Methods in Molecular Biology. 2013; 936:313-24.
67. Liu CJ, Kao SY, Tu HF, et al. Increase of microRNA miR-31 level in plasma could be a potential marker of oral cancer. Oral Diseases. 2010; 16:360-4.

2.

Variations in mucins and mucin–associated carbohydrate antigens expression in mucoepidermoid carcinomas

Johannes H. Matse, Wiresh K. Bharos, Enno C.I. Veerman, Elisabeth Bloemena and Jan G.M. Bolscher

Submitted for publication in Archives of Oral Biology

Abstract

The aberrant expression of mucins and mucin-type carbohydrates has been described in many types of cancer, including mucoepidermoid carcinoma (MEC), a malignant salivary gland tumor. In this study, we examined the aberrant expression patterns of mucins (MUC1, MUC4, MUC5AC and MUC5B), simple mucin-type carbohydrate antigens (Tn, sialyl-Tn and T) and mature carbohydrate antigens (Lewis^a and sulfo-Lewis^a antigens) in MEC originating from the parotid gland, which normally does not secrete mucins.

We conducted an immunohistochemical study to investigate the presence of mucins and carbohydrates in 2 normal parotid glands and 24 MEC samples originating from the parotid gland. The expression levels were compared with respect to the histological grading and with expression levels of normal tissue surrounding the same gland. Furthermore, 24 MEC samples from non-parotid salivary glands were included.

We observed loss of topology of membrane-bound MUC1 and MUC4, and de novo expression of MUC5AC, MUC5B and sialyl-Tn in MEC that originated in the parotid gland. Furthermore, mucins MUC1, MUC4 and carbohydrate antigens Tn, sialyl-Tn, T, Lewis^a and sulfo-Lewis^a were overexpressed in MEC samples compared to surrounding normal salivary gland tissues. MUC1 was expressed in both low- and high grade MECs, whereas MUC4 was not expressed in high grade MECs of the parotid gland.

During the development of MEC in the parotid gland, the genes for gel-forming secretory mucins are switched on. Besides these MEC tissues overexpress short oligosaccharides, suggesting that the glycosylation machinery is impaired.

Introduction

Mucins are a heterogeneous family of large glycoproteins with a great diversity of O-linked carbohydrate side-chains which constitute a major part of the mature mucins[1]. They are expressed by epithelial cells and are either membrane-associated or secreted gel-forming[1-4]. Mucins are the main component of the mucous layer that protects epithelial tissue throughout the body[2-4]. In addition, mucins are involved in the differentiation and renewal of the epithelium, modulation of cell adhesion, immune response and cell signaling[2,3].

Membrane-associated mucins are expressed on the apical borders of normal epithelial cells. In malignant cells of epithelial origin, there is a loss of cell membrane polarity which is associated with the activity of a proliferation and survival program, resulting in the transient repositioning of membrane-bound mucins over the entire cell membrane[5]. Most studies that have reported aberrant expression of mucins in cancer are based on the expression of these membrane-bound mucins, such as MUC1 and MUC4. Overexpression of MUC1 and MUC4 has been observed in gastric, pancreatic, prostate, breast, lung, and salivary gland tumors[6-14].

Secreted gel-forming mucins, such as MUC2, MUC5AC, MUC5B, and MUC6, are the main components of the protective mucous layer covering the sensitive epithelial cells and form a physical barrier between the underlying tissue and the environment, each with its own expression preference for certain organ, tissue and cell type. In neoplastic conditions, however, these secretory mucins are highly expressed in carcinomas of different organs, including gastrointestinal, pancreatic, ovary, breast, lung, and salivary glands[10,11,15-21].

Carbohydrates comprise on weight basis > 90% of the mucins. These are attached to serine or threonine residues in the protein backbone and are synthesized by concerted action of glycosyltransferases[1]. In cancer cells, there are dramatic alterations in the glycosylation machinery, at least partially due to the misfolding of glycosyltransferases[22]. This can lead to mal-functional glycosyltransferases resulting in the expression of short carbohydrates antigens Tn (GalNAcα1-O-Ser/Thr), sialyl-Tn (Sialylα2-6GalNAcα1-O-Ser/Thr) and T (Galβ1-3GalNAcα1-O-Ser/Thr). These so called simple mucin-type carbohydrates, which have a restricted expression pattern in normal salivary glands, are increased in several cancers including breast, prostate, colon and salivary gland tumors[23-29].

One of the most common and most studied malignant salivary gland tumor is the mucoepidermoid carcinoma (MEC). MEC is characterized by mucous, intermediate, epidermoid, clear and/or columnar cells and can be classified as low-, intermediate- or high-grade tumors based on histological parameters such as necrosis, anaplasia, neural invasion, mitoses and cystic growth[30,31]. Low grade MEC has well-formed glandular structures or (micro-) cysts lined by a single layer of mucous cells characterized by a swollen appearance due to the presence of high-molecular weight mucins in their cytoplasm. In high grade tumors, the epidermoid cells predominate, and the tumor has a more solid architecture [32]. Aberrant expression of mucins[10,11] as well as mucin-type carbohydrates[23] have been analyzed in salivary gland MECs in relation to diagnostic and prognostic implications.

However, there has been minor attention given to the different characteristics of the individual salivary glands with respect to mucin expression. While there are no large differences in the expression of membrane bound mucins in all normal salivary glands there is a striking difference between the normal parotid glands and the other salivary glands when comparing the secretory gel-forming mucins. All salivary glands but the parotid gland are made up of serous and mucous acini, the latter being responsible for the secretion of mucins, mainly MUC5B[33,34]. Notably, the parotid gland is a purely serous gland which inherently does not express any secretory gel-forming mucin[35]. Therefore, we have chosen for the present study MEC that originate specifically from the parotid gland to analyze the expression of the membrane-associated and secretory gel-forming mucins as well as some mucin-type carbohydrates. We conducted an immunohistochemical survey to examine the aberrant expression patterns of the membrane-associated MUC1 and MUC4 and the secretory gel-forming mucins MUC5AC and MUC5B. Furthermore the expression levels of the simple mucin-type carbohydrate antigens Tn, sialyl-Tn and T, and the mature carbohydrate antigens Lewis^a (Galβ1-3GlcNAc[Fucα1-4]β1-O-Ser/Thr) and sulfo-Lewis^a (Sulfo-3Galβ1-3GlcNAc[Fucα1-4]β1-O-Ser/Thr) were determined. For comparison reasons minor salivary glands were included in the survey.

We observed a MEC-associated loss of topology intimated by aberrant expression of MUC1, MUC4 and impaired glycosylation of mucin-associated carbohydrate. Strikingly, we observed de novo expression of MUC5AC and MUC5B, as well as the sialyl-Tn antigen in the MEC samples originating from the parotid gland.

Normal and MEC samples

Tumor samples of 48 formalin-fixed-paraffin-embedded (FFPE) MEC samples and 6 normal salivary glands samples (2 parotid, 2 submandibular, and 2 sublingual glands) were retrieved from the archives of the Department of Pathology, VU University medical center (Amsterdam, The Netherlands). All tumors had been surgically removed between 1984 and 2011; in 2 cases only incisional biopsies were taken. HE slides were reviewed by an experienced pathologist (EB). Diagnoses were confirmed, and tumors were graded according to the WHO criteria[36]: 30 low grade, 7 intermediate grade and 11 high grade MEC (Table 1).

Twenty-four tumors were located in the parotid gland and 24 minor salivary glands (11 palate and 13 elsewhere in oral cavity). The design of this study adheres to the code for proper secondary use of human tissue established by the Dutch Federation of Biomedical Scientific Societies[37].

Table 1: Patients Clinicopathological Characteristics.

Clinicopathological Characteristics	Low grade MEC (n=30)	Intermediate grade MEC (n=7)	High grade MEC (n=11)
Mean age in years (range)	48 (9-82)	42 (14-58)	68 (43-82)
Gender			
Male	16	4	7
Female	14	3	4
Tumor location			
Parotid gland	15	5	4
Minor salivary gland	15	2	7
Lymph node metastasis	4	1	3
Resection type			
Radical	18	5	7
Irradical	3	0	3
Non-available	9	2	1
Recurrence	3	1	1
Died of the disease	0	0	1
Mean follow-up in months (range)	35 (1-164)	115 (44-234)	33 (1-84)

Immunohistochemistry of mucins and mucin associated carbohydrate antigens

All tumor and normal samples were fixed in 4% buffered formalin, processed and embedded in paraffin according to routine procedures. From each tissue block, 4 µm thick sections were cut on coated slides and dried overnight at 37° C. The sections were deparaffinized in xylene and rehydrated through graded concentrations of ethanol. Endogenous peroxidase was blocked by 0.3% H_2O_2 in methanol for 30 min at room temperature. Immunohistochemistry was subsequently performed using a panel of antibodies and or lectins directed against protein and carbohydrate epitopes on mucins (Table 2). 3,3-Diaminobezidine (DAB) was used as substrate for the Powervision method (Immunologic, Klinipath). Sections were counterstained with Mayer's hematoxylin, dehydrated and mounted.

Table 2: Antibodies and Lectins.

Antigen	Antibody / Lectin	Dilution	Antigen Retrieval
MUC1	Clone E29 (Dako)	1:100	-
MUC4	Clone 1G8 (Zymed)	1:50	Tris/EDTA pH 8.0
MUC5AC	Novov astera	1:50	Na-Citrate pH 6.0
MUC5B	EU1[1]	1:10	-
Tn antigen	HPA-HRP (Sigma-Aldrich)	1:100	-
sialyl-Tn antigen	sTn 219 (GeneTex)	1:50	Na-Citrate pH 6.0
T antigen	PNA-HRP (Sigma-Adrich)	1:50	-
Lewis[a]	7LE (Abcam)	1:100	-
Sulfo-Lewis[a]	F2[1]	1:10	-

[1]) Veerman et al., 1997, 2003

The results of the immunohistochemistry were evaluated by semi-quantitatively scoring the percentage of positive neoplastic cells by two investigators (JHM and WB) and included the overall tumor slide. Tumors were considered positive when the tumor section of the slide contained 5% or more positive neoplastic cells. Distinction was made between expressions in mucous cells versus non-mucous cells (epidermoid and intermediate cells). The expression of mucins and carbohydrate epitopes in the MEC samples was recorded for tumor part and the surrounding normal tissue separately.

Results

Histology

Normal salivary glands as well as the unaffected tissues surrounding the salivary gland tumors showed the histological appearances that are characteristic for the gland type, i.e. in the parotid gland tissue only serous acini were present (Figures 1 and 2), the sublingual gland tissue for the most part, was composed of mucous acini, and the submandibular gland contained both serous and mucous acini (not shown).

Figure 1: Immunohistochemical staining of MUC1 (A), MUC4 (B), MUC5AC (C), and MUC5B (D), in normal parotid gland and in MEC tissue originating from the parotid gland. In the normal parotid tissue shows epithelial cells lining the striated or excretory ducts (black arrow) and serous acini (S). In the MEC tissues mucous (*) and non-mucous (#) cells are indicated. Magnification 100x.

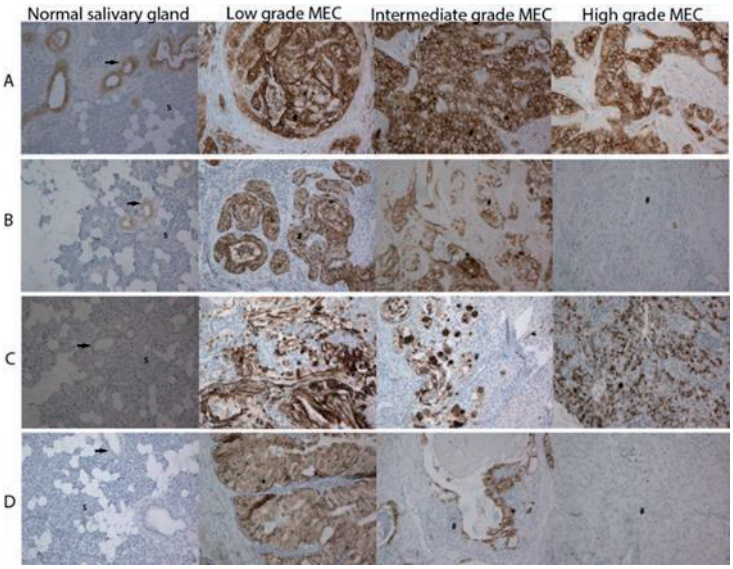
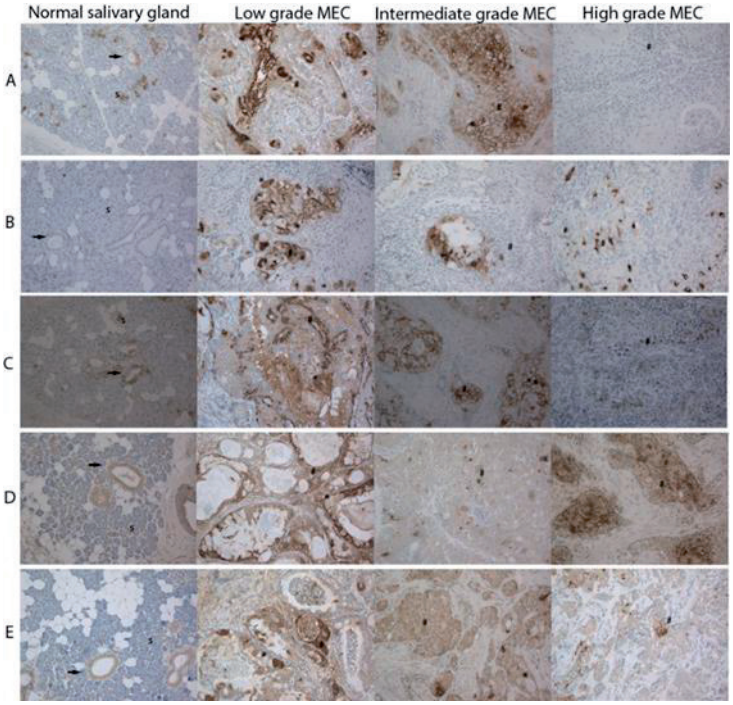


Figure 2: Immunohistochemical staining of simple mucin-type carbohydrate antigens, Tn (A), Sialyl-Tn (B), and T antigen(C) and mature carbohydrate antigens, Lewisa (D), and sulfo-Lewisa (E) in normal parotid gland and in MEC tissue originating from the parotid gland. In the normal parotid tissue shows epithelial cells lining the striated or excretory ducts (black arrow) and serous acini (S). In the MEC tissues mucous (*) and non-mucous (#) cells are indicated. Magnification 100x.



The MEC samples used in the study were histologically consistent with the histological grade as described (Brandwein et al., 2001; Goode & El-Naggar, 2005). Low grade MEC samples showed well-formed glandular structures or (micro-) cysts lined by a single layer of mucous cells, while high grade MEC samples had a more solid architecture. In low grade MEC samples, mucous cells were the predominant cell type. In high grade MEC samples, non-mucous (epidermoid and intermediate) cells were the predominated cell type.

Mucin expression in normal salivary gland tissue

The parotid gland showed expression of the membrane-associated mucins MUC1 and MUC4 in the apical membrane of epithelial cells lining the striated and the excretory ducts. Comparable expression was found in all other salivary gland types. MUC1 and MUC4 were not present in the cytoplasm of serous or mucous acinar cells; the latter cells only focally expressed some MUC4. The secreted salivary mucin, MUC5B, was exclusively located in the cytoplasm of mucous acinar cells of the sublingual and submandibular gland tissue. The purely serous parotid gland, which is devoid of mucous acini, was negative for MUC5B. None of the normal salivary gland tissues showed expression of the gastric mucin MUC5AC. (Data are summarized in Table 3 and typical examples are given in Figure 1).

Table 3: Mucins and mucin-type carbohydrates expression in normal salivary glands.

Normal salivary glands	MUC1	MUC4	MUC5AC	MUC5B	Tn	Sialyl-Tn	T	Lewis[a]	Sulfo-Lewis[a]
Mucous acini[1]	-	+/-	-	+	+	-	+	+	+
Serous acini	-	-	-	-	+/-	-	+/-	-	-
Ducts	+	+	-	-	+	-	+	+	-

[1]) Note, parotid glands contain only serous acini while the other salivary glands contain both serous and mucous acini.

Carbohydrate antigen expression in normal salivary gland tissue

The simple mucin-type carbohydrate antigens Tn- and T were present on the membranes and in the cytoplasm of ductal cells of the normal parotid gland as well as in the mucous acinar cells of all other normal salivary glands. Occasionally, serous acinar cells were also stained for these antigens. Neither normal parotid gland tissue nor the other salivary gland tissues showed Sialyl-Tn expression (parotid gland is shown in Fig. 2). The parotid gland showed expression of the Lewis[a] antigen in the ductal cells but not sulfo-Lewis[a]. The other normal gland tissues showed both antigen, Lewis[a] in ductal cells and mucous acinar cells, while sulfo-Lewis[a] was detected only in mucous acinar cells. Data are summarized in Table 3 and typical examples are given in Figure 2.

Table 4: Mucin and mucin-type carbohydrates in MEC.

MEC tissue	MUC1	MUC4	MUC5AC	MUC5B	Tn	Sialyl-Tn	T	Lewis[a]	Sulfo-Lewis[a]
Parotids gland MEC samples (n=24)									
Mucous cells	22	19	18	20	21	13	16	19	15
Non-mucous cells	23	16	17	18	16	19	20	22	13
Minor gland MEC samples (n=24)									
Mucous cells	21	18	16	16	16	6	14	16	14
Non-mucous cells	21	18	13	10	17	10	15	17	15

Mucin expression in MEC samples

The mucin expression profiles in normal tissues flanking the tumor were similar to those in normal salivary glands. Within the tumorous tissue, however, both mucous and non-mucous cells were positively stained for mucins (Fig. 1 and Table 4). Unlike in normal salivary glands, in which MUC1 and MUC4 are located exclusively in the apical membrane, within the tumor these mucin species were distributed over the entire cell membrane as well as in the cytoplasm. MUC5AC and MUC5B epitopes were detected in the cytoplasm of mucous cells and in the cytoplasm of non-mucous cells of the MEC tissues (Table 4). This is a striking finding since MUC5B is absent from normal parotid glands and MUC5AC is absent in all salivary glands.

Carbohydrate antigen expression in MEC

The expression of carbohydrate antigens in tumor-flanking normal tissue was similar to that of normal salivary gland tissues. In the MEC tissue samples, the simple carbohydrate antigens Tn and T as well as the mature carbohydrate antigens Lewis[a] and sulfo-Lewis[a] were localized on the cell membranes and in the cytoplasm of non-mucous cells. In mucous cells, these carbohydrates were mainly present in the cytoplasm (Fig. 2 and Table 4). The sialyl-Tn antigen, which is absent in normal salivary gland tissues, was found in the cytoplasm of both mucous and non-mucous MEC cells, most prominently so for the non-mucous cells in MEC from the parotid gland. Overall the carbohydrate antigens Tn, sialyl-Tn, T, Lewis[a] and sulfo-Lewis[a] were overexpressed in MEC samples compared to surrounding normal salivary gland tissues.

Differences between low, intermediate and high grade MEC

In Table 5 an overview is given of the expression of the various epitopes in relation to the histological grade of the MECs. It shows that in virtually all low grade MEC samples, mucous as well as non-mucous cells stained positive for all mucin types and carbohydrates tested.

The low grade MECs of the parotid gland and minor glands exhibited a comparable staining pattern. Compared to the low grade MECs, a smaller proportion of the high grade MECs was positive for the mucin and carbohydrate epitopes. MUC1 was the only mucin that was present in virtually all of the high grade MEC samples. MUC4 was absent in the high grade MEC samples from the parotid gland.

Table 5: Mucin and mucin-type carbohydrate in MEC of different grades.

	MUC1	MUC4	MUC5AC	MUC5B	Tn	Sialyl-Tn	T	Lewis[a]	Sulfo-Lewis[a]
Parotid gland MEC samples (n=24)									
Low (n=15)									
Mucous	14	15	14	15	15	11	13	15	11
Non-mucous	15	15	15	12	13	15	15	14	8
Intermediate (n=5)									
Mucous	4	4	3	3	5	2	3	3	3
Non-mucous	4	1	2	4	2	2	4	5	4
High (n=4)									
Mucous	4	0	1	2	1	0	0	1	1
Non-mucous	4	0	0	2	1	2	1	3	1
Minor gland MEC samples (n=24)									
Low (n=15)									
Mucous	13	13	13	12	13	5	10	12	11
Non-mucous	14	13	10	5	12	7	12	13	9
Intermediate (n=2)									
Mucous	2	2	2	2	2	1	1	2	0
Non-mucous	2	2	1	2	2	2	1	1	2
High (n=7)									
Mucous	6	3	1	2	1	0	3	2	3
Non-mucous	5	3	2	3	3	1	2	3	4

Discussion

Previous studies compared the expression of several MUC species in MEC of salivary glands with focus on the diagnostic and prognostic implications[10,11]; in particular MUC1, MUC4, MUC5AC and MUC5B were aberrantly expressed. Because the parotid gland is the only purely serous salivary gland which normally does not secrete gel-forming mucins, we have compared the expression of these mucins in MECs originating from the parotid gland to expressions of mucins in MECs from minor salivary glands. MECs from both types of glands showed essentially comparable expression patterns. MUC1 and MUC4, which in normal tissues have an apical membrane localization, were distributed over the entire membrane in MEC tissues. A similar aberrant distribution has been reported for other cancer types, and highlights the loss of membrane polarity tumors[6-14]. A difference in expression of these membrane-bound mucins was observed, depending on the histological grade of the tumor. In the parotid gland MUC1 was expressed in both low- and high grade MECs, whereas MUC4 was not expressed in the high grade MECs. In the minor glands MECs, this difference in expression was less pronounced.

The salivary mucin MUC5B and the gastric mucin MUC5AC were expressed in low grade MECs from minor glands but also in MECs from the parotid glands. This is remarkable because, under normal conditions, neither of the mucins are present in the parotid gland. Moreover, in the normal minor glands only MUC5B and MUC5AC is rarely detectable. The presence of MUC5B and MUC5AC in 'non-mucous' cells that lack the swollen morphology of mucous cells is intriguing. The histological appearance of mucous cells is thought to be caused by secretory mucins filling the cell-interior after their release from granules during the processing of tissue sections. The fact that mucin-positive, 'non-mucous' cells in MEC tissues lack this characteristic appearance suggests that these mucins are not fully maturated, e.g. due to incomplete multimerization and/or glycosylation. Probably, the secretory MUC-genes are switched-on while the parotid tissue is not prepared to fully process these mucins.

Besides aberrant expression of mucins, also mucin-type carbohydrate antigens were differently expressed in MEC, notably the initiation of the expression of sialyl-Tn in low grade MECs originating from the parotid gland. Sialylation of Tn will terminate the elongation of Tn to larger oligosaccharides, indicating that in MEC (mucin)glycoproteins are decorated with short, immature oligosaccharide sidechains. Impaired glycosylation has been attributed to a somatic mutation in the chaperone protein for T-synthase, Cosmc, resulting in the loss of T-synthase activity because the newly misfolded T-synthase is immediately degraded[38,39].

As examples of mature oligosaccharides, we also compared the expression of Lewis^a and sulfo-Lewis^a. In normal salivary gland, tissue the Lewis^a antigen is present in both ductal and mucous cells, in the latter as part of the carbohydrate moiety of MUC5B. In contrast, sulfo-Lewis^a, which is highly specific for MUC5B mucins, in normal gland tissue only labels mucous acini[35]. In low grade MEC, Lewis^a was present both in mucous and in non-mucous cells, whereas the sulfo-Lewis^a antigen was found primarily in mucous cells and to a lesser

extent in non-mucous cells. This suggests that mucous cells in MEC tissues express mature mucin glycoproteins. On the other hand, it can be envisaged that serous cells are present in MEC tissues that express both sulfo-Lewis^a and MUC5B protein epitopes, yet lack the mucous phenotype.

This study demonstrated MEC-associated abnormal expression of mucins and mucin-type carbohydrates. Notably, the aberrant expression profiles in MEC samples that originated from the parotid gland suggests de novo synthesis of MUC5B and MUC5AC as well as sialyl-Tn rather than overexpression.

Acknowledgements

We would like the acknowledge W. de Jong, Department of Pathology, VU University medical center, Amsterdam for his technical assistance. JGMB and ECIV were supported by a grant from the University of Amsterdam for research into the focal point 'Oral Infections and Inflammation'.

References

1. Varki A, Cummings RD, Esko JD, et al. (2009). Essentials of Glycobiology (2nd ed.). New York: Cold Spring Harbor Laboratory Press, (Chapter 9).
2. Moniaux N, Escande F, Porchet N, et al. Structural organization and classification of the human mucin genes. Frontiers in Bioscience. 2001; 6:1192-206.
3. Hollingsworth MA, & Swanson BJ. Mucins in cancer: Protection and control of the cell surface. Nature Reviews Cancer. 2004; 4:45-60.
4. Corfield AP. Mucins: A biologically relevant glycan barrier in mucosal protection. Biochimica et Biophysica Acta. 2015; 1850:236-52.
5. Kufe D. Mucins in cancer: function, prognosis and therapy. Nature Reviews Cancer. 2009; 9:874-85.
6. Giuntoli RL, Rodriguez GC, & Whitaker RS. Mucin gene expression in ovarian cancers. Cancer Research. 1998; 58:5546-50.
7. Liu B, Lague JR, Nunes DP, et al. Expression of membrane-associated mucins MUC1 and MUC4 in major human salivary glands. Journal of Histochemistry and Cytochemistry. 2002; 50:811-20.
8. Llinares K, Escande F, Aubert S, et al. Diagnostic value of MUC4 immunostaining in distinguishing epithelial mesothelioma and lung adenocarcinoma. Modern Pathology. 2014;17:150-7.
9. Saitou M, Goto M, Horinouchi M, et al. MUC4 expression is a novel prognostic factor in patients with invasive ductal carcinoma of the pancreas. Journal of Clinical Pathology. 2005; 58:845-52.
10. Alos L, Lujan B, Castillo M, et al. (2005). Expression of membrane-bound mucins (MUC1 and MUC4) and secreted mucins (MUC2, MUC5AC, MUC5B, MUC6 and MUC7) in mucoepidermoid carcinomas of salivary glands. American Journal of Surgical Pathology. 2005; 29:806-13.
11. Handra-Luca A, Lamas G, Bertrand JC, et al. MUC1, MUC2, MUC4, and MUC5AC expression in salivary gland mucoepidermoid carcinoma - Diagnostic and prognostic implications. American Journal of Surgical Pathology. 2005; 29:881-9.
12. Singh AP, Chauhan SC, Bafna S, et al. Aberrant expression of transmembrane mucins, MUC1 and MUC4, in human prostate carcinomas. The Prostate. 2006; 66:421-9.
13. Senapati S, Sharma P, Bafna S, et al. The MUC gene family: their role in the diagnosis and prognosis of gastric cancer. Histology & Histopathology. 2008; 23:1541-52.
14. Khodarev NN., Pitroda SP, Beckett MA, et al. MUC1-induced transcriptional programs associated with tumorigenesis predict outcome in breast and lung cancer. Cancer Research. 2009; 69:2833-7.
15. Hanski C, Hofmeier M, Schmitt-Gräff A, et al. Overexpression or ectopic expression of MUC2 is the common property of mucinous carcinomas of the colon, pancreas, breast, and ovary. Journal of Pathology. 2003; 182:385-91.
16. Reis CA, David L, Carvalho F, et al. Immunohistochemical study of the expression of MUC6 mucin and co-expression of other secreted mucins (MUC5AC and MUC2) in human gastric carcinomas. Journal of Histochemistry and Cytochemistry. 2000; 48:377-88.
17. Perrais M, Pigny P, Buisine MP, et al. Aberrant expression of human mucin gene MUC5B in gastric carcinoma and cancer cells. Identification and regulation of a distal promoter. Journal of Biological Chemistry. 2001; 276:15386-96.

18. Sóñora C, Mazal D, Berois N, et al. Immunohistochemical analysis of MUC5B apomucin expression in breast cancer and non-malignant breast tissues. Journal of Histochemistry and Cytochemistry. 2006; 54:289-99.
19. Park SY, Roh SJ, Kim YN, et al. Expression of MUC1, MUC2, MUC5AC and MUC6 in cholangiocarcinoma: Prognostic impact. Oncology Reports. 2009; 22: 649-57.
20. Dijkema T, Terhaard CHJ., Roesink JM, et al. MUC5B levels in submandibular gland saliva of patients treated with radiotherapy for head-and-neck cancer: A pilot study. Radiation Oncology. 2012; 7:91.
21. Li Y, Wang X, Ao M, et al. Aberrant mucin5B expression in lung adenocarcinoma detected by iTRAQ labeling quantitative proteomics and immunohistochemistry. Clinical proteomics. 2013; 10:15.
22. Stowell SR, Ju, T, & Cummings RD. Protein glycosylation in cancer. Annual Review of Pathology. 2015; 10:473-510.
23. Therkildsen MH, Mandel U, Christensen M, et al. Simple mucin-type Tn and sialosyl-Tn carbohydrate antigens in salivary gland carcinomas. Cancer. 1993; 72:1147-54.
24. Therkildsen MH, Mandel U, Thorn J, et al. Simple mucin-type carbohydrate antigens in major salivary gland. Journal of Histochemistry & Cytochemistry. 1994; 42:1252-9.
25. Carneiro F, Santos L, David L, et al. T-(Thomsen-Friedenreich)-Antigen and Other Simple Mucin-Type Carbohydrate Antigens in Precursor Lesions of Gastric-Carcinoma. Histopathology. 1994; 24:105-13.
26. Colpitts TL, Billing P, Granados E, et al. Identification and immunohistochemical characterization of a mucin-like glycoprotein expressed in early stage breast carcinoma. Tumor Biology. 2002; 23:263-78.
27. Nakagoe T, Sawai T, Tuji T, et al. Prognostic value of expression of sialosyl-Tn antigen in colorectal carcinoma and transitional mucosa. Digestive Diseases & Sciences. 2002; 47:322-30.
28. Brockhausen, I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. EMBO Reports. 2006; 7:599-604.
29. Kirkeby S, Moe D, & Bardow A. MUC1 and the simple mucin-type antigens: Tn and sialyl-Tn are differently expressed in salivary gland acini and ducts from the submandibular gland, the vestibular folds, and the soft palate. Archives of Oral Biology. 2010; 55:830-41.
30. Brandwein MS, Ivanov K, Wallace DI, et al. Mucoepidermoid carcinoma - A clinicopathologic study of 80 patients with special reference to histological grading. American Journal of Surgical Pathology. 2001; 25:835-45.
31. Goode RK, & El-Naggar A K. (2005). Mucoepidermoid carcinoma. In: L. Barnes, J. W. Eveson, P. Reichert, & D. Sidransky (Eds), World Health Organization classification of tumours. Pathology and genetics of head and neck tumours pp 219-20. Lyon: IARC Press.
32. Goode RK, Auclair PL, & Ellis GL. Mucoepidermoid carcinoma of the major salivary glands - Clinical and histopathologic analysis of 234 cases with evaluation of grading criteria. Cancer. 1998; 82:1217-24.

33. Bolscher J, Veerman ECI, Van Nieuw Amerongen, A, et al. Distinct populations of high-M(r) mucins secreted by different human salivary-glands discriminated by density-gradient electrophoresis. Biochemical Journal. 1995; 309:801-6.
34. Veerman ECI, Bolscher JGM, Appelmelk BJ, et al. A monoclonal antibody directed against high M(r) salivary mucins recognizes the SO3-3Gal β1-3GlcNAc moiety of sulfo-Lewisa: a histochemical survey of human and rat tissue. Glycobiology. 1997; 7:37-43.
35. Veerman ECI, van den Keijbus PAM, Nazmi K, et al. istinct localization of MUC5B glycoforms in the human salivary glands. Glycobiology. 2003; 13:363-6.
36. Barnes L, Eveson JW, Reichart P, et al. (2005). World Health Organization Classification of Tumours. Pathology and Genetics of Head and Neck Tumours. Lyon: IARC Press, (Chapter 1).
37. Oosterhuis JW, Coebergh JW, & van Veen EB. Tumour banks: well-guarded treasures in the interest of patients. Nature Reviews Cancer. 2003; 3:73-7.
38. Ju TZ, & Cummings RD. Identification of a unique molecular chaperone Cosmc-1 required for activity of the mammalian core 1 beta 3-galactosyltransferase. Proceedings of the National Academy of Sciences of the U.S.A. 2002; 99:16613-8.
39. Aryal RP, Ju TZ, & Cummings RD. The Endoplasmic Reticulum Chaperone Cosmc Directly Promotes in Vitro Folding of T-synthase. Journal of Biological Chemistry. 2010; 285:2456-62.

3.

High number of chromosomal copy number aberrations inversely relates to t(11;19)(q21;p13) translocation status in Mucoepidermoid Carcinoma of the Salivary Glands

Johannes H. Matse, Enno C.I. Veerman, Jan G.M. Bolscher, C. René Leemans, Bauke Ylstra and Elisabeth Bloemena

Submitted for publication in Oncotarget

Abstract

Although rare, mucoepidermoid carcinoma (MEC) is one of the most common malignant salivary gland tumors. The presence of the t(11;19)(q21;p13) translocation in a subset of MECs has raised interest in genomic aberrations in MEC. In the present study we conducted genome-wide copy-number-aberration analysis by micro-array comparative-genomic-hybridization on 27 MEC samples.

Low/intermediate-grade MECs had significantly fewer copy-number-aberrations compared to high-grade MECs (low vs high: 3.48 vs 30;p=0.0025; intermediate vs high: 5.7 vs 34.5;p=0.036). The translocation-negative MECs contained more copy-number-aberrations than translocation-positive MECs (average amount of aberrations 15.9 vs 2.41;p =0.04).

Within all 27 MEC samples, 16p11.2 and several regions on 8q were the most frequently gained regions , while 1q23.3 (RGS4) was the most frequently detected loss.

Low/intermediate-grade MEC samples had copy-number-aberrations in chromosomes 1, 12 and 16, while high-grade MECs had a copy-number-aberration in 8p. The most commonly observed copy-number-aberration was the deletion of 3p14.1 (FOXP1), which was observed in 4 of the translocation-negative MEC samples. No recurrent copy-number-aberrations were found in translocation-positive MEC samples.

Based on these results, we conclude that MECs may be classified as follows: (i) t(11;19)(q21;p13) translocation-positive tumors with no or few chromosomal aberrations and (ii) translocation-negative tumors with multiple chromosomal aberrations

Introduction

Mucoepidermoid carcinoma (MEC), although rare, is the most common malignant salivary gland neoplasm. According to the WHO, MEC can be classified as low-, intermediate- or high-grade tumors based on the histological parameters necrosis, anaplasia, neural invasion, mitoses and percentage cystic growth[1]. Prognosis of high grade MEC is worse than that of low and intermediate grade tumors[2].Determining genomic aberrations within MEC has gained interest because aberrations may be used as a classification tool for MEC. Earlier studies found that a subset of MEC carries a t(11;19)(q21;p13) translocation, leading to the *CRTC1-MAML2* fusion gene[3-8]. MECs that harbor the t(11;19)(q21;p13) translocation generally have a more favorable prognosis than translocation negative tumors , irrespective of histological grade[5,6]. A few studies have investigated the genomic copy number aberrations in MEC using micro-array comparative genomic hybridization (arrayCGH) technique. They have shown that low grade MEC samples, in general had fewer copy number aberrations than high grade MEC samples[7,8]. Furthermore, these studies found that translocation-positive MEC samples had fewer copy number aberrations compared to translocation-negative MEC samples. Both studies reported the loss of 9p21.3 and the gain of 5p15.33 and 8q24.3 regions. Anzick et al[7] found the loss of the 9p21.3, which harbors the CDKN2AB gene, exclusively in translocation-positive MEC samples), whereas Jee et al[8] reported that translocation-negative MEC samples also harbor this genomic aberration. Both studies concluded that the loss of the *CDKN2A/B* genes was associated with an unfavorable prognosis.

Due to the fact that MEC constitutes a group of diverse, non-frequently occurring tumors and considering different copy number aberrations reported in literature, confirmation of these results in another sample set seems to be warranted. Therefore, we have conducted a genomic analysis using arrayCGH to gain insight into chromosomal copy number in MEC. We compared the aberrations with histological grade and translocation status of each sample. Results suggest that two types of MECs can be distinguished: (i) a group of MECs without t(11;19)(q21;p13) translocation with many copy number aberrations (> 6), independent of histological grade, and (ii) a group of MECs with the t(11;19)(q21;p13) translocation with no or a few copy number aberrations (<6) with two exceptions classified as low and intermediate grade.

Materials and Methods

Samples

Formalin-fixed paraffin-embedded MEC samples and matched healthy salivary gland samples were retrieved from the archives of the Department of Pathology, VU University medical center, Amsterdam, The Netherlands. All tumors were surgically removed between 1984 and 2012. Hematoxylin and eosin stained sections (10 µm) were reviewed by an experienced pathologist (EB) who confirmed the original diagnosis and graded the tumors. Twenty-seven cases of which there was no doubt about the diagnosis were used for this study. Clinicopathological details are described in Table 1. The design of the study adhered to the code for proper secondary use of human tissue established by the Dutch Federation of Biomedical Scientific Societies (http://www.federa.org)[9].

Table 1: Clinicopathological details of the 27 MEC samples.

Sample[1]	Sex	Age	Tumor Site[2]	Recurrence/ metastasis	t-status[3]
LG1	f	9	PG	-/+	-
LG2	m	79	MSG	-/-	+
LG3	m	45	PG	-/-	-
LG4	f	24	PG	-/-	+
LG5	f	13	PG	-/-	+
LG6	m	34	MSG	-/-	-
LG7	m	81	PG	-/-	+
LG8	f	42	PG	-/-	+
LG9	m	50	PG	-/-	+
LG10	f	25	MSG	-/-	+
LG11	f	45	PG	-/-	-
LG12	f	64	MSG	+/-	+
LG13	m	30	MSG	-/-	+
LG14	m	57	PG	-/-	+
LG15	m	62	PG	-/-	-
LG16	f	43	PG	-/-	+
LG17	m	51	PG	-/-	-
IntG1	m	24	MSG	-/-	+
IntG2	f	58	PG	-/-	-
IntG3	f	71	MSG	+/-	+
IntG4	m	14	PG	-/-	+
IntG5	m	58	PG	-/-	-
IntG6	f	52	MSG	+/+	+
HG1	m	81	PG	-/-	-
HG2	m	43	PG	-/-	+
HG3	f	82	PG	-/-	+
HG4	f	59	MSG	-/+	-

[1]Histological grading: LG, low grade; IntG, intermediate grade; HG, high grade

[2]Tumor site: PG, parotid gland; MSG, minor salivary gland

[3]Presence of translocation t(11;19)(q21;p13)

DNA isolation

DNA was isolated as previously described[10]. Briefly, 6 sections of 10 μm were deparaffinized, macro-dissected and incubated with 1M sodium thiocyanate at 38°C, for 16 h, followed by a proteinase K treatment at 55°C for another 16 h. DNA was isolated using the QIAmp DNA micro-kit (Qiagen, Hilden, Germany). Purity and quantity of the DNA samples was measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, USA).

ArrayCGH

ArrayCGH was performed as described previously[11]. Although FFPE is not the most ideal material for aCGH analyses, we have over the years built a large amount of experience herewith generating good quality data[11]. Equal amounts (500 ng) of DNA from MEC samples and from matching normal salivary gland tissue were labeled with cyanine 3'-dUTP (Cy3) and cyanine 5'-dUTP (Cy5) nucleotides (Enzo Life Sciences, Farmingdale, NY, USA). Free nucleotides were removed using the MinElute PCR Purification Kit (Qiagen). Oligonucleotide arrayCGH was performed using the SurePrint G3 Human CGH Microarray Kit, containing 180880 *in situ* synthesized 60-mer oligonucleotides representing 169793 unique chromosomal locations evenly distributed over the genome (space ~17kb) and 4548 additional unique oligonucleotides, located at 238 of the Cancer Census genes (4x180K array, Agilent Technologies, Palo Alto, CA, USA). The exact array design can be found online in the Gene Expression Omnibus (GEO) GPL8687 (http://www.ncbi.nlm.nih.gov/geo).). The data are accessible through GEO number GSE87353. Segment values were converted to calls by setting thresholds corresponding to 20% of the tumor cells with that copy number aberration: this percentage converts to a log2 ratio of > 0.1375 for gains and < -0.1520 for losses and all values in between are called normal copy number. Values above the 0.1375 threshold were called gains, values below the -0.1520 threshold were called losses. Although these settings are relatively liberal, this threshold was nevertheless chosen for the optimal balance between background and detection of real copy number aberrations. Nevertheless, for analyses, we have focused on recurrent copy number aberrations that occurred in at least three tumors. The log2 ratio threshold for high copy number amplification and homozygous deletion were 1.0 and -1.0, respectively. The data were analyzed using Nexus.

To make sure that the detected copy number alterations were real and not background static, we considered a copy number real when it was present in 3 or more MEC samples.

FISH analysis

For detection of the translocation in MEC samples, fluorescence *in situ* hybridization (FISH) analysis was carried out on 4 μm tissue sections according to the manufacturer's protocol, using ZytoLight ® SPEC MAML2 Dual Color Break Apart Probe (ZytoVision Ltd, Bremerhaven, Germany). The nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI), diluted in Vectashield, and samples were evaluated by fluorescence microscopy (ZyGreen: excitation 503nm, emission 528nm; ZyOrange: excitation 547nm, emission 572nm). Cells without the

t(11;19)(q21;p13) translocation show fused green and red signals, typically resulting in a yellow signal. Translocation-positive cells exhibit fused green and red, as well as separated red and green signals, or split signal (Figure 1). A MEC sample was considered positive for the t(11;19)(q21;p13) translocation when the split signal was identified in at least 10 out of 100 cells.

Statistical analysis

Differences between the presence of copy number aberrations in low grade, intermediate grade and high grade MEC, translocation-positive and translocation-negative MEC samples were determined using the Mann-Whitney test. A two-sided P-value of ≤ 0.05 was considered to be statistically significant. Statistical analyses were performed with the use of the Statistical Package for the Social Sciences (SPSS version 20.0).

Results

Clinicopathological Characteristics

The clinical and histopathological characteristics of patients and tumors in the study are shown in Table 1. The mean age of patients was 48 years (range 9-82). Eighteen tumors originated in the parotid gland and 9 tumors originated in minor salivary gland. Three patients had loco-regional recurrence of the tumor and 3 other patients developed a metastasis. One patient died of the disease. Of the 27 MECs, 17 were classified as low grade, 6 as intermediate grade and 4 as high grade. FISH analysis revealed that 17 of the total 27 MEC samples harbored the t(11;19)(q21;p13) translocation. These comprised 11 of the 17 low grade, 4 of the 6 intermediate grade and 2 of the 4 high grade samples (Table 1).

Genomic Profiles in MEC

ArrayCGH profiles of 27 MECs are presented in Figure 1. To be sure that we were picking up real copy number aberrations and not static we considered a copy number aberration real when it was being found in at least 3 samples). Using this criteria we found 37 gain and 23 losses (for all copy number alterations see Table 2). The most common copy number aberration was the loss of the 1q23.3 region, which was found in 5 MEC samples.

Figure 1: Genome-wide frequency plot (top). The Y-axix represents the percentage of the total group (n=27). Underneath the genome-wide frequency plot, the individual arrayCGH profiles of the of 27 MEC samples (17 low; 6 intermediate; 4 high grade) used in the study. Gains are portrait in blue and losses are portrait in red.

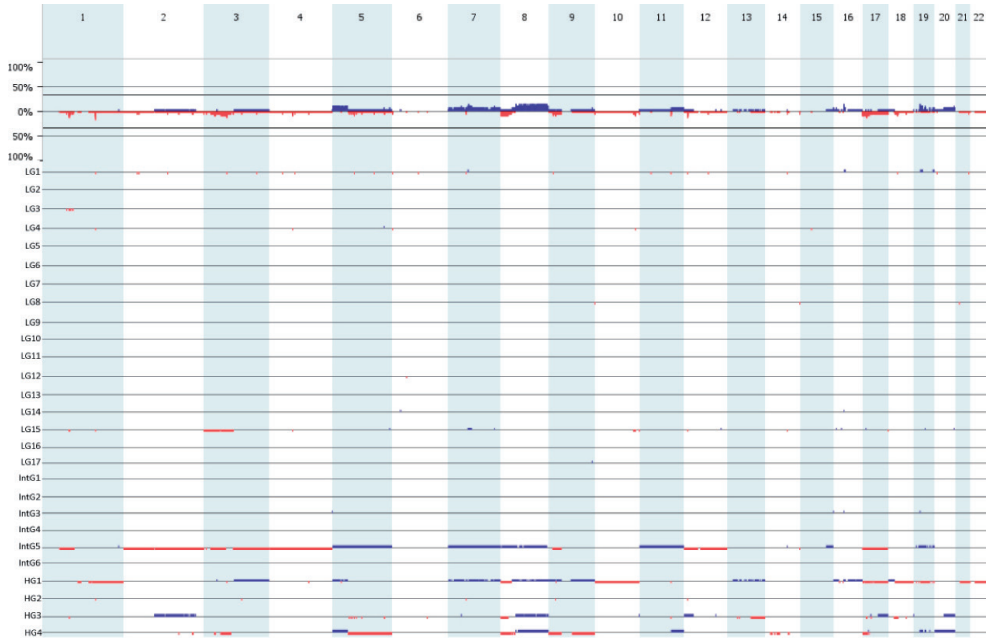


Table 2: Recurring copy number aberrations in 27 MEC samples.

Chromosome band	Region coordinates	Region Length (bp)	Event	No. of samples	Candidate genes*
1p31.1	81271921-82317822	1045901	Loss	3	ADGRL2
1p31.1	82317822-83535674	1217852	Loss	4	
1p31.1-p22.3	83535674-85092491	1556817	Loss	3	
1q23.3	162789883-163066609	276726	Loss	5	RGS4
3p23-p22.2	31951947-38114334	6162387	Loss	3	MiR-26a-1, CTDSPL
3p21.1-p14.1	51942767-69002771	17060004	Loss	3	PBRM1, ADAMTS9
3p14.1	71138660-71270186	131526	Loss	4	FOXP1
5p15.33 - p14.2	0-24691408	24691408	Gain	3	PDCD6 TRIO
5p14.1	26738646-26959771	221125	Gain	3	
5p14.1-p13.3	26959771-33389493	6429722	Gain	3	DROSHA
5p13.3-p13.2	33389493-36093914	2704421	Gain	3	ADAMTS12, TARS, RAD1
5p13.1-q11.1	38884395-47700000	8815605	Gain	3	DAB2
5q12.3-q13.1	66116648-67826992	1710344	Loss	3	PIK3R1
7p14.1	38969086-40708290	1739204	Gain	3	POU6F2
7q11.1-q11.21	59100000-62420609	3320609	Gain	3	
7q11.21-q11.23	62420609-71886202	9465593	Gain	3	AUTS2
7q34	140141708-140795070	653362	Gain	3	BRAF
8p23.3-p21.2	0-24154814	24154814	Loss	3	DEFA1, DEFB1, DLC1, MTUS1
8q11.1-q11.21	45200000-48882980	3682980	Gain	4	
8q11.21-q11.22	48882980-51942456	3059476	Gain	3	SNAI2
8q12.1	52901077-56253836	3352759	Gain	3	
8q12.1	56253836-57369226	1115390	Gain	4	PLAG1
8q12.1	57369226-59262605	-1893379	Gain	3	
8q12.1-q12.2	59262605-61867310	2604705	Gain	4	CYP7A1, SDCBP
8q12.2-q12.3	61867310-63949108	2081798	Gain	3	
8q12.3-q13.1	63949108-66631815	2682707	Gain	4	CYP7B1
8q13.1-q13.2	66631815-69704074	3072259	Gain	3	
8q13.2-q13.3	69704074-71870918	2166844	Gain	4	SULF1
8q13.3-q21.12	71870918-79997004	8126086	Gain	4	

8q21.12-q21.13	79997004-82856098	2859094	Gain	3	FABP5
8q21.13-q21.2	82856098-85047238	2191140	Gain	4	
8q21.2-q21.3	85047238-87786949	2739711	Gain	3	WWP1
8q21.3	87786949-90946365	3159416	Gain	4	
8q21.3	90946365-92054675	1108310	Gain	3	
8q21.3-q22.3	92054675-102317166	10262491	Gain	4	RUNX1T1, CDH17, TP53INP1
8q22.3	102317166-104094744	1777578	Gain	3	
8q22.3-q23.1	104094744-107148474	3053730	Gain	4	BAALC, CTHRC1
8q23.1	107939667-109558687	1619020	Gain	3	ANGPT1
8q23.1-q24.23	109558687-139487231	29928544	Gain	4	HAS2, TNFRSF11B, MYC
8q24.23-q24.3	139487231-140865305	1378074	Gain	3	TRAPPC9
8q24.3	140865305-146274826	5409521	Gain	4	PTK2, MAFA, MAPK15
9p23-p22.3	13398392-14364288	965896	Loss	3	
9p21.3	20451378-20600059	148681	Loss	3	
9p21.3	21851680-23259683	1408003	Loss	3	CDKN2A, CDKN2B, MTAP
11q21	95420121-95735922	315801	Loss	3	MAML2
12p13.2	11187958-11860096	672138	Loss	4	ETV6, PRB1
16p11.2	31753818-33345523	1591705	Gain	4	
17p13.3	0-197784	197784	Loss	3	
17p13.3-p13.1	927102-7097922	6170820	Loss	3	PLD2, MiR134
17p13.1-p12	7771326-11457982	3686656	Loss	3	ALOX15B
17p12	11457982-12504361	1046379	Loss	4	MAP2K4
17p12	12504361-13787450	1283089	Loss	3	ELAC2
17p12	13787450-14819919	1032469	Loss	4	
17p12-p11.2	14819919-17694955	2875036	Loss	3	
17p11.2-q11.1	19523822-22200000	2676178	Loss	3	MAP2K3
18q12.1	26913332-28779980	1866648	Loss	3	DSC3
19p12	19860022-20054862	194840	Gain	3	
19p12-q11	20054862-28500000	8445138	Gain	3	
19q12	34912078-36065910	1153832	Gain	3	
20q11.1-q13.33	27100000-59628718	32528718	Gain	3	WISP2, MAFB, MIR296

*gene location according to UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly.

Analysis based on histological grading (Figure 2) revealed that there was no differences between the number of copy number alterations between low and intermediate grade MEC samples (p=0.763) and between intermediate and high grade MEC samples (p=0.099). Therefore, we combined the low and intermediate grade MEC samples to form a low/intermediate grade MEC group.

Low/intermediate grade MEC samples had statistically significantly fewer copy number aberrations compared to high grade MEC samples (mean values: low/intermediate grade vs high grade: 3.48 vs 30, p = 0.0025). The deletion of 1p31.1 (containing *ADGRL2*), 1p31.1-p22.3 and 12p13.2 (containing *ETV6*), and the gain of 16p11.2 were exclusively found in the low/intermediate grade MEC samples, while the loss of 8p23.3-p12 (containing *DEFB1*, *DLC1*, *MTUS1*) was exclusively found in high grade MEC samples (Table 3).

Figure 2: Genome-wide comparison between low (n=17) vs. intermediate (n=6), intermediate (n=6) vs. high (n=4), and low (n=17) vs. high (n=17) grade MEC samples. Gains are portrait in blue and losses are portrait in red. Regions of p<0.05 are marked by horizontal bars of gains (blue) and losses (red) on the significance track.

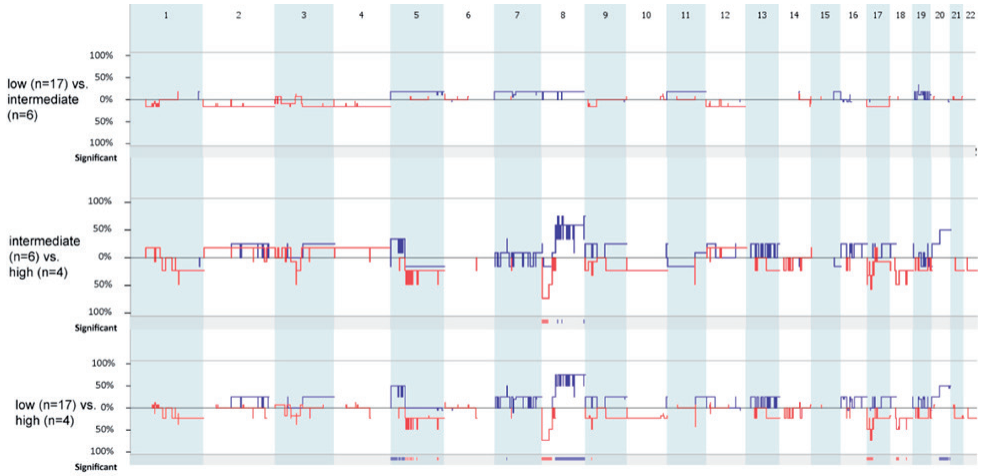


Table 3: Recurrent copy number aberrations exclusively found in low/intermediate grade (n=23) or in high grade MEC samples (n=4).

Chromosome band	Region coordinates	Region Length (bp)	Event	No. of samples	Present in LG/ IntG or HG[1]	Candidate genes*
1p31.1	81271921-82317822	1045901	Loss	3	LG/IntG	ADGRL2
1p31.1-p22.3	83535674-85092491	1556817	Loss	3	LG/IntG	
8p23.3-p21.2	0-24154814	24154814	Loss	3	HG	DEFB1, DLC1, MTUS1
12p13.2	11187958-11860096	672138	Loss	3	LG/IntG	ETV6
16p11.2	31753818-33345523	1591705	Gain	3	LG/IntG	

[1] LG/IntG, low/intermediate grade MEC; HG, high grade MEC.

* gene location according to UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly.

Translocation-positive MEC contained fewer copy number aberrations than translocation-negative tumors (Figure 3) (mean values translocation-positive vs translocation-negative: 2.41 vs 15.9, p = 0.04). In total 22 copy number aberrations were found (11 loss and 11 gains), which were all found to be exclusive for translocation-negative MEC samples (see Table 4 for all copy number aberrations that were present in 3 or more translocation-negative MEC samples). The most common copy number aberration detected was the loss of 3p14.1 (containing *FOXP1*), which was found in 4 translocation-negative MEC samples (LG1, LG15, HG1 and HG4).

Figure 3: Genome-wide comparison between t(11;19)(q21;p13) translocation-positive (n=17) and t(11;19)(q21;p13) translocation-negative MEC samples (n=10). Gains are portrait in blue and losses are portrait in red. Regions of p<0.05 are marked by horizontal bars of gains (blue) and losses (red) on the significance track.

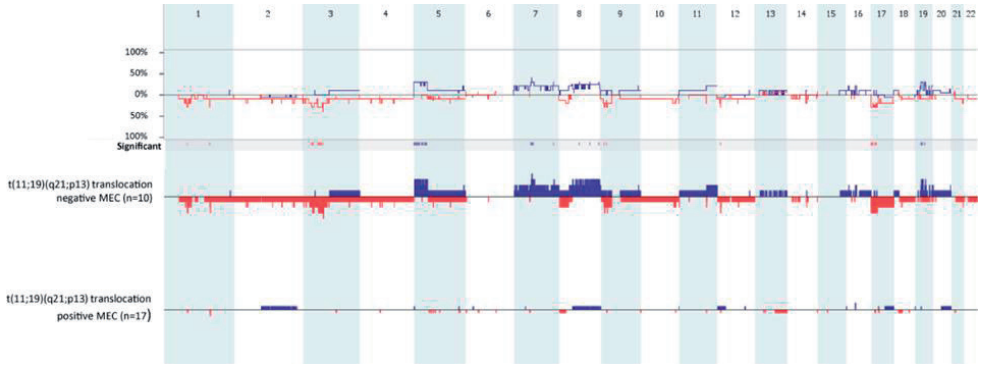


Table 4: Recurring copy number aberrations exclusively found in MEC samples without the t(11;19)(q21;p13) translocation (n=10).

Chromosome band	Region coordinates	Region length (bp)	Event	No. of samples	Candidate gene(s)*
1p31.1	81271921-82317822	1045901	Loss	3	ADGRL2
3p23-p22.2	31951947-38114334	6162387	Loss	3	MiR-26a-1, CTDSPL
3p14.1	71138660-71270186	131526	Loss	4	FOXP1
5p15.33 - p14.2	0-24691408	24691408	Gain	3	PDCD6, TRIO
5p14.1	26738646-26959771	221125	Gain	3	
5p14.1-p13.3	26959771-33389493	6429722	Gain	3	DROSHA
5p13.3-p13.2	33389493-36093914	2704421	Gain	3	ADAMTS12, TARS, RAD1
5p13.1-q11.1	38884395-47700000	8815605	Gain	3	DAB2
7q11.1-q11.21	59,100,000-62,420,609	3320609	Gain	3	
7q11.21-q11.23	62420609-71886202	9465593	Gain	3	AUTS2
7q34	140141708-140795070	653362	Gain	3	BRAF
8q24.23-q24.3	139487231-140865305	1378074	Gain	3	TRAPPC9
9p23-p22.3	13398392-14364288	965896	Loss	3	
9p21.3	21851680-23259683	1408003	Loss	3	CDKN2A, CDKN2B, MTAP
17p13.3-p13.1	927102-7097922	6170820	Loss	3	PLD2, MiR134
17p13.1-p12	7771326-11457982	3686656	Loss	3	ALOX15B,
17p12	11457982-12504361	1046379	Loss	3	MAP2K4
17p12	12504361-13787450	1283089	Loss	3	ELAC2,
17p12-p11.2	14819919-17694955	2875036	Loss	3	
17p11.2-q11.1	19523822-22200000	2676178	Loss	3	MAP2K3
19p12-q11	20054862-28500000	8445138	Gain	3	
19q12	34912078-36065910	1153832	Gain	3	

*gene location according to UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly.

Discussion

Mucoepidermoid carcinomas are salivary gland tumors with a variable histopathological differentiation. They have an unpredictable clinical outcome, which poses significant diagnostic and therapeutic challenges. Analysis of genomic aberrations may help in the classification of these tumors, but large scale analysis of the genomic imbalance in MEC is hampered because of its rather low frequency of occurrence. There are several arrayCGH based studies described as yet[7,8,12], using different spatial resolution, with different cut-offs for specific genetic aberration for all or for a subset of MEC. To us, this warranted further analysis of the genomic imbalance in an additional set of 27 well documented MECs.

In the present study, copy number aberrations were found in 14 out of 27 MEC samples (Table1). Based on the amount of copy number aberrations, two groups of MEC can be distinguished. One group with no or few copy number aberrations (6 or less), and another with multiple copy number aberrations (19 or more); the latter being about 22% of the total number of MEC samples. MEC samples harboring multiple copy number aberrations were found amongst all three histological grades.

With the exception of one sample (HG3), most of the MEC with multiple copy number aberrations were translocation-negative, while translocation-positive MEC samples had mostly 6 or less copy number aberrations. These findings confirmed the consensus in literature and may explain why t(11;19)(q21;p13) translocation-positive MEC have a more favorable survival outcome compared to transformation negative MECs, which are characterized by chromosomal instability[7,8,12].

Comparison with previous studies[7,8,12] underscores that MEC is extremely diverse with respect to copy number aberrations. For instance, the most frequently detected loss in the present set of MECs, 1q23.3 (containing *RGS4*) in 5 of the 27 MECs (Table 2), was not reported in the other studies, whereas the most frequently detected loss reported by Jee et al[8], 18q12.2-qter, was found only once in our sample set. Similarly, the majority of most frequently detected gains found in this study (Table 2) did not correspond with earlier research[8]. Only a set of small regions that were gained in chromosome 8 corresponded with a larger gain 8q11.1-q12.2 described by Jee et al[8].

Only one region found in our study, the lost region 9p21.3, was recorded to be by earlier studies[7,8,12]. The loss of 9p21.3, containing *CDNK2A/B,* initially suggested this occurrence to be specific to translocation-positive MEC[7], but later was also found in translocation-negative MEC[8]. In both studies, the loss of *CDKN2A/B* was associated with an unfavorable prognosis. Zhang et al[12] found that the region 9p21.3 was also deleted in adenoid cystic carcinoma and salivary duct carcinoma. Furthermore, the deletion of 9p21.3 is a frequent oncogenic event observed in head and neck squamous cell carcinomas, and in lung cancer[13-15].

The regions 8q23.1-q24.23 (myelocytomatosis oncogene (*MYC*)) and 8q24.3 (protein tyrosine kinase 2 (*PTK2*)) were gained in 4 samples, which all had multiple copy number aberrations. *PTK2* has been shown to be gained in prostate, gastric, colorectal cancers and

in salivary adenoid cystic carcinoma[16-19]. *MYC* is overexpressed in nearly 50% of all human tumors[20,21], resulting in the aberrant expression of *MYC* target genes. The t(11;19)(q21;p13) fusion protein can bind and activate *MYC,* leading to cellular transformation via functional complementation of CREB and *MYC* transcription networks[22,23]. In our study however, only one of the 4 samples with a gain of 8q23.1-q24.23 harbored the t(11;19)(q21;p13) translocation.

Although some chromosomal aberrations may suggest the involvement of certain genes, the general instability of these malignancies may also be of importance for MEC development. Therefore, clinically, the instability itself should be taken as a marker rather than specific (onco)genes that are gained or lost in specific samples.

In conclusion, in this study we showed that salivary gland MEC may be classified as follows: (i) MEC with no or few chromosomal aberrations, which are in general positive for the t(11;19)(q21;p13) translocation, and (ii) MEC with multiple genomic imbalances, which are in general t(11;190(q21;p13) translocation negative. This implies that there are different oncogenic pathways within MEC, in which either the fusion-gene or the loss of genetic instability plays a role in the underlining pathologic process.

References

1. Goode RK, El-Naggar AK. 2005 Mucoepidermoid carcinoma. In: WHO Classification of Tumours. Pathology and Genetics of Head and Neck Tumours. Ed. Barnes, L, Eveson JW, Reichart P., Sidransky D. IARC Press.
2. Katabi N, Ghossein R, Ali Set al. Prognostic features in mucoepidermoid carcinoma of major salivary glands with emphasis on tumour histologic grading. Histopathology. 2014; 65:793-804.
3. Nordkvist A, Gustafsson H, Juberg-Ode M, et al . Recurrent rearrangements of 11q14-22 in mucoepidermoid carcinoma. Cancer Genetics and Cytogenetics. 1994; 74:77-83.
4. Tonon G, Modi S, Wu L, et al. t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. Nature Genetics. 2003; 33:208-13.
5. Behboudi A, Enlund F, Winnes M, et al. Molecular classification of mucoepidermoid carcinomas - prognostic significance of the MECT1-MAML2 fusion oncogene. Genes Chromosomes and Cancer 2006; 45:470-81.
6. Okabe M, Miyabe S, Nagatsuka H, et al. MECT1-MAML2 fusion transcript defines a favorable subset of mucoepidermoid carcinoma. Clinical Cancer Research. 2006; 12:3902-7.
7. Anzick SL, Chen WD, Park Yet al. Unfavorable prognosis of CRTC1-MAML2 positive mucoepidermoid tumors with CDKN2A deletions. Genes Chromosomes and Cancer. 2010; 49:59-69.
8. Jee KJ, Persson M, Heikinheimo K, et al. Genomic profiles and CRTC1-MAML2 fusion distinguish different subtypes of mucoepidermoid carcinoma. Modern Pathology 2013; 26:213-22.
9. Oosterhuis JW, Coebergh JW, van Veen EB. Tumour banks: well-guarded treasures in the interest of patients. Nature Reviews Cancer 2003; 3:73-7.
10. van Essen HF, Ylstra B. High-resolution copy number profiling by array CGH using DNA isolated from formalin-fixed, paraffin-embedded tissues. Methods in Molecular Biology. 2012; 838:329-41.
11. Krijgsman O, Israeli D, van Essen HF, et al. Detection limits of DNA copy number alterations in heterogeneous cell populations. Cellular Oncology. 2013; 36:27-36
12. Zhang L, Mitani Y, Caulin C, et al. Detailed genome-wide SNP analysis of major salivary carcinomas localized subtype-specific chromosome site and oncogenes of potential clinical significance. American Jounal of Pathology. 2013; 182:2048-2057.
13. Rocco JW, Sidransky D. p16(MTS-1/CDKN2/INK4a) in cancer progression. Experimental Cell Research. 2001; 264:42-55.
14. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nature Review Cancer. 2006; 6:107-16.
15. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. The New England Journal of Medicine. 2008; 358:1118-28.
16. Saramäki OR, Porkka KP, Vessella RL, et al. 2006. Genetic aberrations in prostate cancer by microarray analysis. International Journal of Cancer. 2006; 119:1322-9.
17. Killian A, Di Fiore F, Le Pessot F, et al. A simple method for the routine detection of somaticquantitative genetic alterations in colorectal cancer. Gastroenterology. 2007; 132:645-53.

18. Park JH, Lee BL, Yoon J, et al. Focal adhesion kinase (FAK) gene amplification and its clinical implications in gastric cancer. Human Pathology. 2010; 41:1664-73.
19. Chen D, Zhang B, Kang J, et al. Expression and clinical significance of FAK, ILK, and PTEN in salivary adenoid cystic carcinoma. Acta Oto-Laryngologica. 2013; 133:203-8.
20. Nilsson JA, Cleveland JL. Myc pathways provoking cell suicide and cancer. Oncogene. 2003; 22:9007-21.
21. Meyer N, Penn LZ. Reflecting on 25 years with MYC. Nature Review Cancer. 2008; 8:976- 90.
22. Amelio AL, Fallahi M, Schaub FX, et al. CRTC1/MAML2 gain-of-function interactions with MYC create a gene signature predictive of cancers with CREB-MYC involvement. Proceedings of the National Academy of Sciences. 2014; 111:E3260-8
23. Wu L, Liu J, Gao P, et al. Transforming activity of MECT1-MAML2 fusion oncoprotein is mediated by constitutive CREB activation. The EMBO Journal. 2005; 24:2391-402.

4.

Human salivary microRNA in patients with parotid salivary gland neoplasms

Johannes H. Matse, Janice Yoshizawa, Xiaoyan Wang, David Elashoff, Jan G.M. Bolscher, Enno C.I. Veerman, C. René Leemans, D. Michiel Pegtel, David T.W. Wong and Elisabeth Bloemena

Published in: PLoS One, 2015; 10:e0142264

Abstract

Currently, clinical examination, ultrasound scanning (with or without fine needle aspiration cytology), preoperative CT-scan and MRI are available for the differential diagnosis of parotid gland swelling. A preliminary non-invasive salivary diagnostic tool may be helpful in the clinical decision making process. Altered salivary micro-RNA (miRNA) expression levels have been observed in saliva from patients with various cancers. Therefore, we investigated miRNA expression levels in saliva samples from patients with a parotid gland neoplasm using Human miRNA cards in comparison to controls.

In the discovery phase, eight miRNAs were identified having different expression levels in patients compared to controls. In the validation phase, the differences in miRNA expression levels between patients and controls were confirmed for seven out of eight discovered miRNAs ($p < 0.001$). A combination of two miRNAs yielded a receiver-operator-characteristics curve with an AUC of 0.94 (95% CI: 0.87-1.00; sensitivity 91%; specificity 86%). Validation of discovered miRNAs in segregated collected parotid saliva revealed that expression of these miRNAs differ between whole saliva and parotid saliva.

A two miRNA combination can predict the presence of a parotid gland neoplasm. Furthermore, this study suggested that the identified, patient-specific, salivary miRNAs were not derived from the parotid gland itself.

Introduction

MicroRNAs (miRNAs) are small non-coding RNAs consisting of 19-25 nucleotides. They are involved in various cellular processes such as cell differentiation, proliferation and survival. MiRNAs inhibit translation by binding to complementary sequences in the 3' UTR of multiple target mRNAs, usually resulting in their silencing[1,2]. Altered levels of miRNA expression have been implicated in the etiology of cancer[3,4].

Many studies have demonstrated the differential expression of miRNAs between cancer and the related normal surrounding tissue. Specific miRNA expression profiles were identified in a number of cancers[1,5-9]. Fluid-specific miRNA expression profiles have also been identified [1,10,11]. The expression of hsa-miR-21, hsa-miR-141 and the hsa-miR-200 family were up-regulated in serum from patients with ovarian cancer[12]. Furthermore, differences in miRNA expression profiles were found in cystic fluid of intraductal papillary mucinous neoplasm of the pancreas and in urine samples from patients with urinary bladder cancer. In saliva samples from patients with esophageal cancer, the expression level of three miRNAs (hsa-miR-10b*, hsa-miR-144, and hsa-miR-451) was significantly higher compared to the control group. Changes in miRNA expression levels have also been described in saliva samples from patients with oral squamous cell carcinoma[7,13-15]. The expression level of two miRNAs (hsa-miR-125a and hsa-miR-200a) was significantly lower in saliva from patients than in healthy control subjects[7]. Recently, we identified 6 potential biomarkers (mmu-miR140-5p, hsa-let-7g, hsa-miR-15b, hsa-miR-132, hsa-miR-222, and hsa-miR-374) that were differently expressed in whole saliva from patients with benign parotid neoplasms and malignant parotid neoplasms[16]. A combination of four miRNAs (hsa-miR-15b, hsa-miR-132, hsa-miR-223 and mmu-miR-140-5p) did discriminate between patients with a malignant tumor and those with a benign salivary gland tumor (sensitivity 69%, specificity 95%).

For clinical decision making, it is important to determine whether salivary gland swelling is due to a neoplasm or a reactive process. Nowadays, clinical examination and invasive methods such as ultrasound scanning (with or without fine needle aspiration cytology), preoperative CT-scan and MRI are the most common diagnostic methods used[17].

In the present study, we have investigated whether salivary miRNA expression profiles of whole saliva from patients with a parotid gland neoplasm differ from those of healthy controls. We found that the expression levels of seven validated miRNAs were higher in whole saliva from patients compared to controls. Using a combination of two of these, it was possible to discriminate between patients with a parotid gland neoplasm and controls. Furthermore, we found differences in the expression of the validated miRNAs between whole saliva and parotid saliva from the same patient.

Materials and Methods

Patients and salivary samples

Discovery phase

For the discovery phase, unstimulated whole saliva samples from 20 patients with a parotid gland neoplasm, obtained from the Salivary Gland Tumor Biorepository (SGTB) at the MD Anderson Cancer Clinic in Houston, TX, USA were tested. Data with regard to the miRNA expression in this group has already been described[16]. Ten unstimulated whole saliva samples from healthy controls were obtained from the Samsung Medical Center in Seoul, South-Korea (Table 1). No clinical information other than age, gender, ethnicity and tumor subtype was available.

Validation in whole saliva

The miRNAs emerging from the discovery phase were validated in an independent sample set of whole saliva. This set consisted of 32 randomly selected unstimulated whole saliva samples from patients with a parotid gland neoplasm and was obtained from the SGTB; 14 were collected at the Department of Otolaryngology and Head-Neck Surgery, VU University medical center, Amsterdam, The Netherlands.

Unstimulated whole saliva samples from 14 healthy controls were collected at the Department of Oral Biochemistry, Academic Centre for Dentistry Amsterdam, Amsterdam, The Netherlands (Table 1).

Validation in parotid saliva

Besides whole saliva, we also collected parotid saliva from 12 patients with a parotid gland neoplasm. Parotid saliva samples were collected separately from both the affected and the contralateral (normal) parotid gland using salivettes (Sarstedt, Nümbrecht, Germany) at the VU University medical center. As a control, parotid saliva was collected from 14 healthy controls. All parotid saliva was collected upon stimulation with 2% citric acid (Table 1).

Table 1: Patients' characteristics of neoplasm samples and control samples used in the discovery and the validation phases.

Discovery phase in Whole saliva (n=30)			Validation phase in Whole saliva (n=60)			Validation phase in Parotid saliva (n=26)		
	Neoplasm	**Control**		**Neoplasm**	**Control**		**Neoplasm**	**Control**
Mean age (range)	56.5 (33-82)	52.4 (39-67)	Mean age (range)	55.0 (20-89)	47.3 (23-70)	Mean age (range)	56.3 (55-90)	56.3 (55-90)
sex (m/f)	8/12	4/6	sex (m/f)	26/20	7/7	sex (m/f)	5/7	7/7
Ethnicity			Ethnicity			Ethnicity		
Hispanic	2	0	Hispanic	2	0	Hispanic	0	0
Caucasian	16	0	Caucasian	42	10	Caucasian	12	10
Black	2	0	Black	2	1	Black	0	1
Asian	0	10	Asian	0	3	Asian	0	3
Neoplasm subtypes			**Neoplasm subtypes**			**Neoplasm subtypes**		
Pleomorphic adenoma	6		Pleomorphic adenoma	19		Pleomorphic adenoma	7	
Warthins tumor	3		Warthins tumor	6		Warthins tumor	2	
Salivary duct carcinoma	2		Salivary duct carcinoma	6		Salivary duct carcinoma	3	
Mucoepidermoid carcinoma	1		Mucoepidermoid carcinoma	3				
Acinic cell carcinoma	1		Acinic cell carcinoma	3				
Adenoid cystic carcinoma	1		Adenoid cystic carcinoma	2				
Oncocytic carcinoma	1		Adeno-carcinoma, nos	2				
Neuroendocine carcinoma	3		Basal cell adenocarcinoma	1				
Myoepithelial carcinoma	1		Basal cell adenoma	1				
Cystadeno-carcinoma	1		Myoepithelial carcinoma	1				
			Carcinoma ex-pleomorphic adenoma	1				
			Undifferntiated carcinoma	1				

Salivary miRNA profiling

Discovery phase

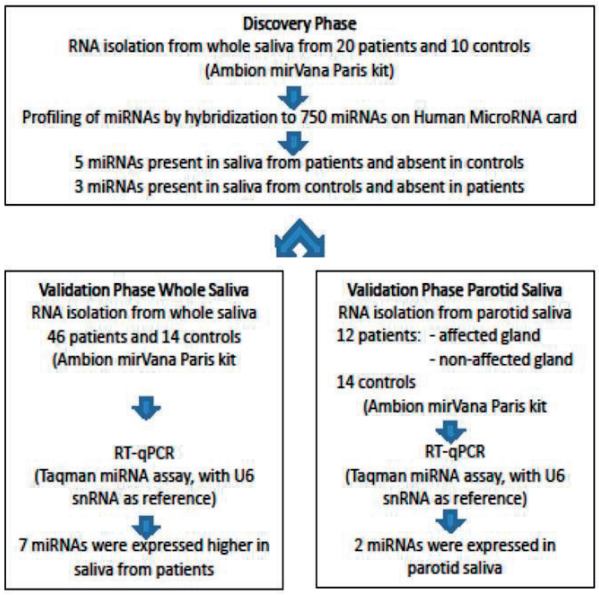
MiRNA profiling of saliva samples from our previous study[16] was re-evaluated in comparison to the miRNA expression levels in saliva from healthy controls. In short, saliva samples were thawed and centrifuged. The cell-free supernatant (cleared saliva) was collected. Total RNA was isolated from cleared saliva using RNA extraction kits (Ambion mirVana Paris kit). Extracted RNA (1-10 ng) was reverse transcribed and pre-amplified. Undiluted pre-amplification product was mixed with 2x Taqman Master mix and loaded onto the Taqman Human MicroRNA cards (Applied Biosystems, Foster City CA). The cards with 750 miRNA were centrifuged and run on the Applied Biosystems 7900HT Fast Real-Time PCR instrument.

The Ct value is defined as the cycle number in the fluorescence emission which exceeds that of a fixed threshold. A Ct of 15-30 was considered high expression, and a Ct of 35 was considered low expression. A Ct value above 40 was considered as undetectable miRNA. We calculated ΔCt by subtracting the Ct value of the reference gene, RNA polymerase III transcribed U6 snRNA [18], from the Ct value of each candidate biomarker. Data normalization was performed using RQ manager 1.2.1 and Data Assist v3.0 from Applied Biosystems. The qPCR based gene expression values between the two groups were compared using the non-parametric Wilcoxon test. Potential miRNA genes were then selected based on $P < 0.05$. Figure 1 gives an overview of research design and saliva samples used in the discovery and validation phase.

Validation phase in whole saliva

Eight miRNAs which showed the highest difference in expression between the samples from patients with a parotid tumor and normal controls and were overlapping between malignant and benign neoplasms were selected for further testing in the validation phase. The selected miRNAs were validated by RT-qPCR using Taqman miRNA assays in an independent sample set consisting of 46 saliva samples from patients with a parotid salivary gland tumor and 14 saliva samples from controls. RT-qPCR was performed as previously described [16]. Briefly, total RNA was reverse-transcribed and pre-amplified. RT-qPCR was done using specific probes for selected miRNAs and was carried out in a 384 well plate using the Roche LightCycler 480 II (Roche, San Francisco, CA). All RT-qPCRs were performed in duplicate for all candidate miRNAs. U6 snRNA was used as a reference miRNA.

Figure 1: Overview of the research design and saliva samples used in the discovery and validation phase.



Validation of discovered miRNA in parotid saliva

RT-qPCR was used to analyze parotid saliva for the presence of the validated miRNAs. Total RNA was isolated from parotid saliva as described above. Parotid saliva samples were analyzed for the presence of miRNAs using Taqman microRNA assays as described above for validation. All qPCRs were performed in duplicate for all candidate miRNA and negative controls (in which RNA was omitted). For miRNA qPCR experiments, U6 snRNA was used as the reference miRNA.

Data analysis

In the discovery phase, miRNA profiling data was analyzed for miRNAs that were absent and/or present in whole saliva from patients with a parotid gland neoplasm in comparison to controls. Wilcoxon rank-sum test was performed to compare the expression of miRNAs in samples from patients with a parotid gland neoplasm to controls in order to determine the level of significance for the miRNA biomarkers. The miRNAs emerging from the discovery phase were validated in an independent set of 60 saliva samples from 46 patients with a parotid gland neoplasm and 14 controls. Wilcoxon rank-sum test was used to compare the ΔCt values for miRNAs between the groups. A two-sided p-value less than 0.05 was deemed statistically significant. Next, multivariate logistic regression analysis was used to construct a classification model to discriminate between patients with a tumor and healthy controls. Backward-stepwise model selection criterion was used to obtain a final model. Specifically, we started fitting a prediction model with all the miRNA biomarkers that differentially

expressed between the two groups. In each model reduction step, the miRNA that was the least significant was removed. We continued by successively re-fitting reduced models and applying the same rule until all remaining miRNAs were statistically significant. ROC curves were constructed to determine the diagnostic/predictive values of individual and combined biomarkers from the logistic model. Area under the curve was computed via numerical integration of the ROC curves. The performance of the model for classification was assessed by identifying the cut-off point of the predicted probability which yielded the largest sum of sensitivity and specificity. Leave-one-out cross-validation was used to validate the logistic regression model and the cross-validation error rate was reported. All statistical analyses were performed by statistical software R Version 2.15.0 (http://www.r-project.org/).

Ethics statement

Permission for this study was obtained from the Medical Ethical Committee of the VU University medical center (registration number: 2009/58), for patients and healthy controls from VU university medical center and the Academic Centre for Dentistry Amsterdam. Also, permission was obtained from the SGTB at MD Anderson Cancer Clinic (IRB: LAB07-0383), the Samsung Medical Center (SMC IRB file #: 2008-01-028-016) and the UCLA Institutional Review Board (UCLA IRB) (ref number: FWA00004642). All participants were adults (> 18 yrs of age) and gave their written informed consent.

Results

Discovery of salivary miRNA markers

In the discovery phase, miRNA expression profiles in whole saliva from patients with a parotid gland neoplasm (both benign and malignant) and miRNA expression profiles in whole saliva from controls were established using Taqman Human MicroRNA cards. 742 miRNAs were analyzed (S1 Table). Of these, 5 miRNAs (hsa-miR-296-5p, hsa-miR-577, hsa-miR-1233, hsa-miR-1267, and hsa-miR-1825) had a significantly higher expression level (lower ΔCt) in whole saliva samples from patients with a parotid gland neoplasm compared to their expression levels in whole saliva from healthy controls. In addition, 3 miRNAs (hsa-miR-103a-3p, hsa-miR-211 and hsa-miR-425-5p) had a significantly higher expression level in whole saliva from controls compared to patients with a parotid gland neoplasm (Table 2).

Table 2: Mean ΔCt (sd) and Wilcoxon 2-sided p-values of discovered salivary miRNA biomarkers. The discovery sample set consisted of whole saliva samples from patients with a parotid gland neoplasm (n = 20) and whole saliva samples from healthy controls (n = 10). ND: non-detectable, assigned to samples with a Ct-value of 40 and higher.

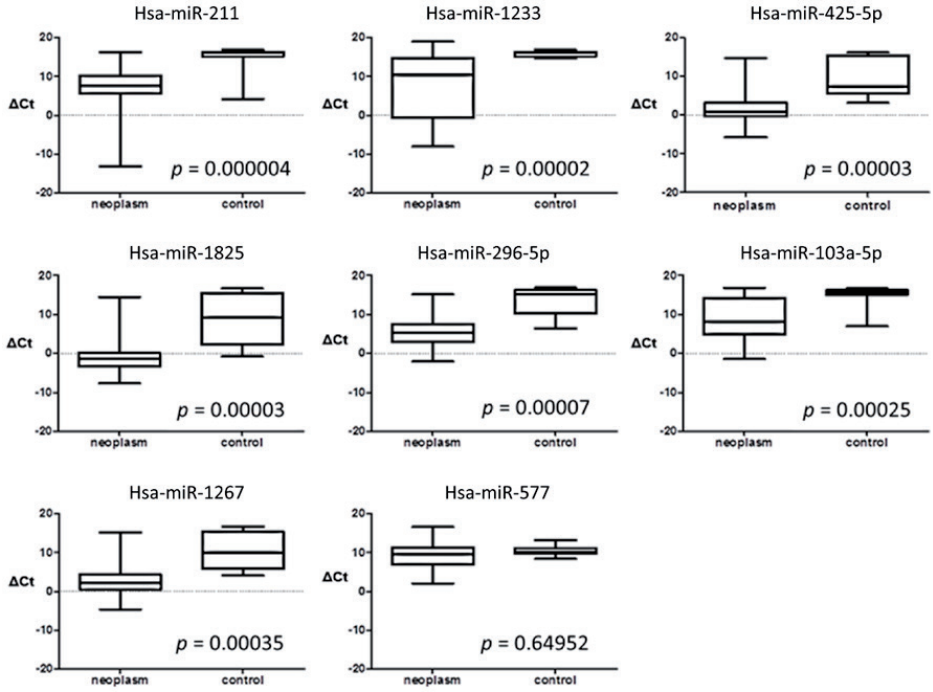
miRNA	ΔCt control mean (sd)	ΔCt neoplasm mean (sd)	Wilcoxon 2-sided p-value
hsa-miR-103a-3p	6.34 (1.71)	ND	< 0.001
hsa-miR-211	9.45 (1.38)	ND	< 0.001
hsa-miR-425-5p	11.56 (3.72)	ND	< 0.001
hsa-miR-296-5p	ND	7.41 (3.75)	< 0.001
hsa-miR-577	ND	5.96 (3.44)	< 0.001
hsa-miR-1233	19.14 (4.98)	8.93 (8.49)	0.002
hsa-miR-1267	ND	3.87 (2.79)	< 0.001
hsa-miR-1825	20.94 (2.26)	5.17 (5.38)	< 0.001

Validation of salivary miRNA markers

The expression levels of the 8 miRNAs identified in the discovery phase were validated by RT-qPCR in a separate independent sample set consisting of whole saliva samples from 60 individuals: 46 patients with a parotid gland neoplasm and 14 healthy controls (S2 Table). Four of the five tumor-specific miRNAs (hsa-miR-296-5p, hsa-miR-1233, hsa-miR-1267, and hsa-miR-1825) had a significantly higher expression level in whole saliva from patients than in whole saliva from healthy controls (Table 3). Hsa-miR-103a-3p, hsa-miR-211 and hsa-miR-425-5p, which in the discovery phase were not expressed in saliva from patients, were in the validation phase expressed in saliva from both controls and patients. The expression levels of hsa-miR-103a-3p, hsa-miR-211 and hsa-miR-425-5p were significantly higher in saliva from

patients than that from controls (Figure 2 and Table 3). In both the control and tumor groups, some of the whole saliva samples were positive for the investigated miRNAs while others were negative, resulting in a large variation in miRNA expression levels.

Figure 2: Box-and-whisker plot of the validation of miRNA expression in whole saliva samples. Samples were collected from patients with a parotid gland neoplasm (n = 46) and controls (n = 14). Whiskers represent maximum and minimum ΔCt values. * p < 0. 001.



Evaluation of validated miRNA biomarkers

Multivariate logistic regression models with these miRNA biomarkers were constructed for the classification of patient samples into parotid gland neoplasm and healthy categories. The initial full model consisted of the seven significant miRNAs from Table 3. We then used backward elimination to identify the most parsimonious model. Receiver operating characteristic (ROC) curves were constructed for various possible cut-points of the predicted probability to obtain another measure of the predictive ability of the fitted models. Areas under the curve (AUC) and their 95% confidence interval were computed. The performance of the models were assessed by identifying the cut-off point of the predicted probability which yielded the largest sum of sensitivity and specificity. The full model with 7 miRNA markers (hsa-miR-103a-3p, hsa-miR-211, hsa-miR-296-5p, hsa-miR-425-5p, hsa-miR-1233,

hsa-miR-1267 and hsa-miR-1825) yielded an AUC of 0.95 (95% CI: 0.88-1.00), a sensitivity of 93% and a specificity of 86%. On the other hand, the reduced model with 2 miRNAs (hsa-miR-211 and hsa-miR-1233) yielded an AUC of 0.94 (95% CI: 0.87-1.00) a sensitivity of 91% and a specificity of 86%, which provided satisfactory prediction as compared to the full model (Figure 3A). We also performed leave-one-out cross validation on the reduced model. The error rate of leave-one-out cross validation was 0.18.

Figure 3: Receiver-operating-characteristics (ROC) curve and dot-plot showing the "predicted probability of neoplasm". A. ROC curves computed from the full (dotted line) and the reduced (solid line) logistic regression models. The full model included 7 miRNAs (hsa-miR-103a-3p, hsa-miR-211, hsa-miR-296-5p, hsa-miR-425-5p, hsa-miR-1233, hsa-miR-1267 and hsa-miR-1825). It yielded an AUC of 0.95 (95% CI: 0.88-1.00), a sensitivity of 93% and a specificity of 86%. The reduced final model included hsa-miR-2 and hsa-miR-1233, which in combination provided satisfactory prediction while being parsimonious. The AUC was 0.94 (95% CI: 0.87-1.00), the specificity was 86% and sensitivity was 91%. B. Predicted probability of neoplasm was based on validation data of 2 validated miRNAs (hsa-miR-211 and hsa-miR-1233). Whole saliva samples from patients with a parotid gland neoplasm had a high "predicted probability of neoplasm", while whole saliva samples from controls with non-neoplastic parotid glands in average had a lower "predicted probability of neoplasm". Lower line represents minimum "predicted probability of neoplasm", middle line represents the mean "predicted probability of neoplasm", and the upper line represents the maximum "predicted probability of neoplasm".

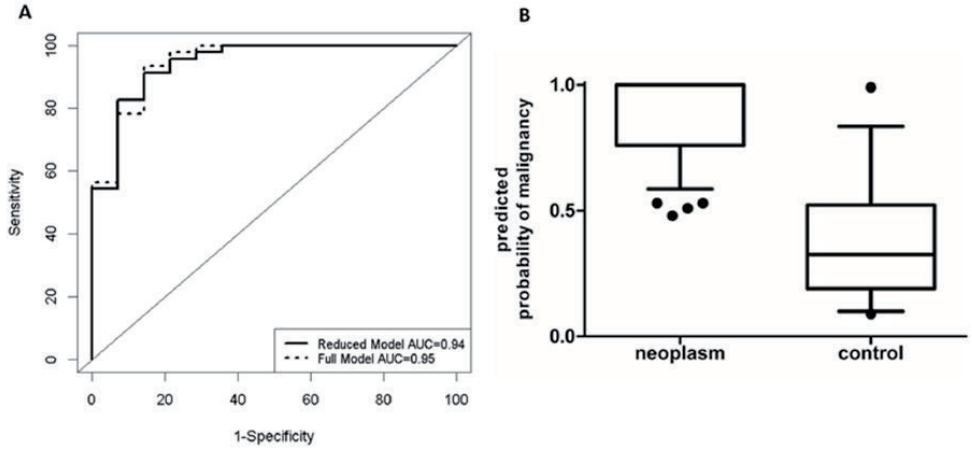


Table 3: Mean ΔCt (sd) and Wilcoxon 2-sided p-values of validated salivary miRNA biomarkers. The independent validation sample set consisted of whole saliva samples from patients with a parotid gland neoplasm (n = 46) and whole saliva samples from healthy controls (n = 14).

miRNA	ΔCt control mean(sd)	ΔCt neoplasm mean (sd)	Wilcoxon 2-sided p-value
hsa-miR-103a-3p	14.32 (3.59)	10.00 (4.75)	< 0.001
hsa-miR-211	14.91 (3.18)	8.56 (4.37)	< 0.001
hsa-miR-296-5p	12.80 (4.15)	6.95 (4.66)	< 0.001
hsa-miR-425-5p	9.96 (5.13)	2.54 (4.79)	< 0.001
hsa-miR-577	10.19 (1.57)	9.66 (2.85)	0.650
hsa-miR-1233	15.74 (0.70)	9.17 (6.89)	< 0.001
hsa-miR-1267	9.78 (5.03)	4.39 (5.21)	< 0.001
hsa-miR-1825	7.47 (7.46)	-0.77 (3.95)	< 0.001

A box-and-whisker plot predicting the probability of neoplasm (Figure 3B) was constructed based on the validation data of a 2 miRNA combination (hsa-miR-1233 and hsa-miR-211). Most samples of the parotid gland neoplasm group scored a probability of neoplasm of 0.8-1.0. In contrast, the predicting probability of neoplasm of the control samples were spread out, ranging from a probability of neoplasm of 0.09 to 0.99 (Figure 3B).

The expression of validated miRNAs in parotid saliva

Next, we investigated whether the 7 validated neoplasm-associated miRNAs in whole saliva from both patients and controls were derived from the affected parotid gland. From 12 patients, saliva from both the affected as well as the healthy parotid gland were collected separately. Similarly, saliva from both parotid glands was collected from 14 healthy controls. Whole saliva samples from these patients and controls had already been analyzed in the validation phase. Two of the 7 validated miRNAs (hsa-miR-425-5p and hsa-miR-1825) were expressed in parotid saliva. The expression levels of these miRNAs differed only marginally between parotid saliva of patients and controls, or, within patients, between saliva from the affected and the non-affected parotid glands. However, the expression level of miRNAs in parotid saliva samples from patients was much lower than in their corresponding whole saliva samples (Table 4). The expression level of hsa-miR-425-p and hsa-miR-1825 in parotid saliva from healthy controls was similar to the expression level of hsa-miR-425-5p and hsa-miR-1825 in whole saliva from healthy controls.

Table 4. Mean ΔCt (sd) of validated salivary miRNA biomarkers in whole saliva and parotid saliva from patients and controls.

	Mean ΔCt (sd) in saliva from patients (n=12)			**Mean ΔCt (sd) in saliva from controls (n=14)**		**p-value**				
miRNA	Affected parotid saliva	non-affected parotid saliva	whole saliva	whole saliva	parotid saliva	affected vs non-affected parotid saliva	affected vs control parotid saliva	non-affected vs control parotid saliva	affected vs control whole saliva	patients whole vs control whole saliva
hsa-miR-103a-3p	15.25 (0.72)	14.46 (0.99)	13.33 (2.87)	14.92 (2.79)	14.32 (3.59)	0.04*	0.63	0.18	0.38	0.03*
hsa-miR-211	15.09 (0.53)	15.39 (0.95)	10.31 (4.18)	14.48 (2.90)	14.91 (3.18)	0.31	0.7	0.37	0.06	<0.001*
hsa-miR-296-5p	14.04 (2.71)	13.96 (3.21)	9.43 (4.92)	13.97 (2.63)	12.80 (4.15)	0.67	0.98	0.63	0.98	0.02*
hsa-miR-425-5p	8.29 (4.44)	9.61 (3.83)	4.23 (4.21)	10.56 (5.43)	9.96 (5.12)	0.34	0.18	0.59	0.32	0.01*
hsa-miR-1233	15.24 (0.72)	15.39 (0.95)	11.54 (7.09)	15.30 (0.76)	15.74 (0.70)	0.54	0.98	0.37	0.09	0.01*
hsa-miR-1267	14.18 (0.97)	14.33 (1.07)	7.24 (4.27)	13.61 (5.03)	9.78 (5.03)	0.62	0.28	0.49	0.16	0.2
hsa-miR-1825	5.87 (3.79)	6.51 (3.45)	0.06 (1.51)	6.26 (1.51)	7.47 (4.66)	0.62	0.55	0.29	0.4	<0.001*

Discussion

In this study, we investigated and validated differences of miRNA expression levels in saliva from patients with a parotid gland neoplasm and healthy controls. The expression of miRNA in whole saliva from patients with a parotid gland neoplasm differs from the expression of miRNA in whole saliva from controls. The expression level of 7 miRNAs was higher in saliva from patients compared to controls and could be confirmed in an independent sample set. A combination of two of these miRNAs (hsa-miR-1233 and hsa-miR-211) enabled a predictive value with a sensitivity of 91%, and a specificity of 86% for predicting the presence of a parotid gland neoplasm.

Today, the invasive fine-needle aspiration cytology (FNAC) technique is one of the most commonly used techniques for the diagnosis of a swelling of the salivary gland. A systematic review with regard to the performance of FNAC in parotid gland lesions concluded that FNAC had a specificity of 97% and a sensitivity of 80%. However the performance variability was relatively large[19]. Therefore, additional (molecular) markers, such as those identified in the present study, can add to the accuracy of a pre-operative diagnosis by FNAC. One of these molecular markers can be salivary miRNA. MiRNAs are stably expressed in all body fluids[2], and are also less prone to being degraded, unlike mRNA or proteins. The ease of saliva sampling, the stable expression of miRNAs and the fact that miRNAs are less prone to be degraded make salivary miRNAs a good choice for biomarkers[20,21].

Strikingly, we observed differences in the expression of miRNA between the discovery and validation phases. In the discovery phase, hsa-miR-103a-3p, hsa-miR-211 and hsa-miR-425-5p were expressed higher in whole saliva samples from controls compared to whole saliva samples from patients. In contrast, these miRNAs had a higher expression level in whole saliva from patients compared to controls in the validation phase. Such differences in miRNA expression have been shown before, highlighting the importance and necessity of validating biomarkers in an independent sample set[22]. Inconsistency between findings of different study groups may be due to a large variation in expression levels of miRNA in controls and patients, or possibly due to variation in miRNA expressions between different ethnic groups as our samples were derived from various sources in both phases of the study. Moreover, we used different histological types of parotid neoplasms, both benign and malignant, which can also be a source of variation, although the relative frequency of the neoplasms was comparable between the discovery and validation phase.

Because miRNAs are known to have a gene regulatory role, we conducted an unbiased search (www. miRBase.org) as to whether the discovered miRNAs had validated targets. Three of the discovered miRNAs (hsa-miR-211, hsa-miR-296-5p, and hsa-miR-425-5p) had validated targets. It has been shown that hsa-miR-211 targets tumor suppressor cadherin 5, which promotes colorectal cancer cell growth in vitro and in vivo[23]. In oral carcinomas, the elevated expression of hsa-miR-211 is associated with poor prognosis[24]. One of the validated targets for hsa-miR-296-5p is the B-cell lymphoma 2 gene, (BCL2) a member of the BCL2-family of

apoptosis regulator proteins. A high expression of hsa-miR-296-5p may result in suppression of BCL2 transcription which, in turn, could lead to higher rates of cell death. One of the many targets of hsa-miR-425-5p is Dicer1, a RNase endonuclease III enzyme that is necessary for miRNA maturation. A study describing the expression of Dicer1 in mucoepidermoid carcinoma found that Dicer1 was expressed higher in MEC compared to surrounding normal tissue [25]. The targets for hsa-miR-211, hsa-miR-296-5p and hsa-miR-425-5p have only been validated in tissue and not in saliva itself. Therefore, we can only speculate what role hsa-miR-211, hsa-miR-296-5p and hsa-miR-425-5p play in saliva.

The present study revealed that only 2 of the 7 validated miRNAs (hsa-miR-425-5p and hsa-miR-1825) were expressed in parotid saliva from patients and controls. The expression of hsa-miR-103a-3p, hsa-miR-211, hsa-miR-296-5p, hsa-miR-1233 and hsa-miR-1267 in whole saliva and the absence of these miRNAs in parotid saliva suggest that these miRNAs are not specifically expressed and/or secreted from the parotid gland tumor. The differences in miRNA expression which we observed between unstimulated whole saliva (Table 3) and stimulated parotid saliva (Table 4) suggest that parotid saliva may have a totally different miRNA expression profile. Other studies have shown that tumors express miRNAs that may be involved in intercellular crosstalk[26,27]. This fits with the hypothesis in which tumor cells transport genetic material, including miRNA, via micro-vesicles (exosomes) to neighboring and/or distant cells, thereby affecting the miRNA expression of those cells and supporting cell growth and progression[9,28]. The miRNAs which are present in blood may leak from the gingival crevice into the oral cavity. This provides whole saliva with many, if not most, of the same molecules found in the systemic circulation including miRNAs.

The initial data presented in this study are a first step toward developing a clinical application to distinguish parotid salivary gland tumors from healthy parotid glands. Future studies should be directed to further unravel the differences between whole saliva and parotid saliva. Although whole saliva is more easily accessible, we cannot conclude, at this moment, which source gives more robust data for tumor diagnostics. Moreover, in the future, differences in miRNA expression in different histological tumor types should also be taken into consideration. Our results justify and urge further investigation towards the discovery of potentially relevant miRNAs in saliva.

References

1. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences of the U.S.A. 2008; 105:10513-18.
2. Weber JA, Baxter DH, Zhang S, et al. The microRNA spectrum in 12 body fluids. Clin Chem. 2010; 56:1733-41.
3. Chen PS, Su JL, Hung MC. Dysregulation of MicroRNAs in cancer. Journal of Biomedical Sciences. 2012; 19:90.
4. Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. Journal of Clinical Oncology. 2009; 27:5848-56.
5. Heneghan HM, Miller N, Kerin MJ. Circulating miRNA signatures: promising prognostic tools for cancer. Journal of Clinical Oncology. 2010; 28:573-4.
6. Nugent M, Miller N, Kerin MJ. MicroRNAs in colorectal cancer: function, dysregulation and potential as novel biomarkers. European Journal of Surgical Oncology. 2011; 37:649-54.
7. Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009; 15:5473-7.
8. Ueda T, Volinia S, Okumura H, et al. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncology. 2010; 11:136-46.
9. Zhang L, Farrell JJ, Zhou H, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology. 2010; 138:949-57.
10. Matthaei H, Wylie D, Lloyd MB, et al. miRNA biomarkers in cyst fluid augment the diagnosis and management of pancreatic cysts. Clinical Cancer Research. 2012; 18:4713-24.
11. Sapre N, Selth LA. Circulating MicroRNAs as Biomarkers of Prostate Cancer: The State of Play. Prostate Cancer. 2013; doi: 10.1155/2013/539680.
12. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. Gynecologic Oncology. 2008; 110:13-21.
13. Brinkmann O, Wong DT. Salivary transcriptome biomarkers in oral squamous cell cancer detection. Advances in Clinical Chemistry. 2011; 55:21-34.
14. Hu S, Arellano M, Boontheung P, et al. Salivary proteomics for oral cancer biomarker discovery. Clinical Cancer Research. 2008; 14:6246-52.
15. Yoshizawa JM, Wong DT. Salivary microRNAs and oral cancer detection. Methods in Molecular Biology. 2013; 936:313-24.
16. Matse JH, Yoshizawa J, Wang X, et al. Discovery and pre-validation of salivary extracellular microRNA biomarkers panel for the non-invasive detection of benign and malignant parotid gland tumors. Clinical Cancer Research. 2013; 19:3032-8.
17. Seifert G, Brocheriou C, Cardesa A, et al. WHO International Histological Classification of Tumours. Tentative Histological Classification of Salivary Gland Tumours. Pathology, Research and Practice. 1990; 186:555-81.

18. Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009; 15:5473-7.
19 Schmidt RL, Hall BJ, Wilson AR, et al. A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions. American Journal of Clinical Pathology. 2011; 136:45-59.
20. Streckfus CF, Dubinsky WP. Proteomic analysis of saliva for cancer diagnosis. Expert Rev Proteomics. 2007; 4:329-32.
21. Thomadaki K, Helmerhorst EJ, Tian N, et al. Whole-saliva proteolysis and its impact on salivary diagnostics. Journal of Dental Research. 2011; 90:1325-30.
22. Pepe MS, Feng Z, Janes H, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. Journal of the National Cancer Institute. 2008; 100:1432-8.
23. Chang KW, Liu CJ, Chu TH, et al. Association between high miR-211 microRNA expression and the poor prognosis of oral carcinoma. Journal of Dental Research. 2008; 87:1063-8.
24. Cai C, Ashktorab H, Pang X, et al. MicroRNA-211 expression promotes colorectal cancer cell growth in vitro and in vivo by targeting tumor suppressor CHD5. PLoS One. 2012; 7:e29750.
25. Chiosea SI, Barnes EL, Lai SY, et al. Mucoepidermoid carcinoma of upper aerodigestive tract: clinicopathologic study of 78 cases with immunohistochemical analysis of Dicer expression. Virchows Archives. 2008; 452:629-35.
26. Kosaka N, Iguchi H, Yoshioka Y, et al. Secretory mechanisms and intercellular transfer of microRNAs in living cells. Journal of Biological Chemistry. 2010; 285:17442-52.
27. Valadi H, Ekström K, Bossios A, et al. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature Cell Biology. 2007; 9:654-9.
28. Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Medicinal Research Reviews. 2012; 32:326-48.

5.

Discovery and pre–validation of salivary extracellular microRNA biomarkers panel for the non–invasive detection of benign and malignant parotid gland tumors

Johannes H, Matse, Janice Yoshizawa, Xiaoyan Wang, David Elashoff, Jan G.M. Bolscher, Enno C.I. Veerman, Elisabeth Bloemena and David T.W. Wong

Published in: Clinical Cancer Research, 2013; 19:3032-8

Abstract

This study is to explore differences in salivary microRNA (miRNA) profiles between patients with malignant or benign parotid gland tumors as a potential pre-operative diagnostic tool of tumors in the salivary glands.

Whole saliva samples from patients with malignant (n=38) or benign (n=29) parotid gland tumors were obtained from the Salivary Gland Tumor Biorepository (SGTB). After total RNA isolation, human miRNA cards were used for miRNA-profiling. The differential miRNA expression was analyzed using 2-sided Wilcoxon test. Real Time-quantitative PCR (RT-qPCR) was used to validate selected miRNAs in an independent sample set. Receiver-operating-characteristics (ROC) curve and probability of malignancy was exploited to evaluate the diagnostic power of the validated miRNAs.

With miRNA-profiling, 57 of 750 investigated miRNAs were differently expressed, of which 54 showed higher miRNA expression in samples from patients with malignant tumors than those from patients with benign tumors. Validating the expression in an independent sample set of 9 miRNAs revealed indeed higher expression of miRNAs in malignant samples compared to benign samples. The expression of 6 validated miRNAs was statistically significantly different between the two groups (P <0.05). A 4-miRNA-combination was able to discriminate between saliva samples from patients with malignant tumors from those of patients with benign parotid gland tumors (sensitivity 69%, specificity 95%).

Salivary miRNA profiles differ in saliva from patients with malignant from saliva from patients with a benign parotid gland tumor. These results are promising to develop a non-invasive diagnostic tool for diagnosing tumors in the salivary glands.

Introduction

Tumors of the salivary glands comprise a diverse group of benign and malignant neoplastic lesions. The World Health Organization distinguishes 37 subtypes of salivary gland tumors**:** 13 benign and 24 malignant salivary gland tumors which constitute approximately 1-7% of all head and neck malignancies and 0.3% of all human tumors. Classification of a tumor in the salivary glands depends upon its histopathological characteristics. Tumors in the salivary glands are challenging to diagnose due to their rarity (0.4-13.5/100,000 cases), the many tumor subtypes and because the tumor subtypes have (partly) overlapping histopathology. Of the three salivary glands in the oral cavity, the parotid gland is the most commonly affected site. Eighty percent of salivary gland tumors occurs in this gland, whereas only 10-15% occurs in the submandibular gland and 5% in the sublingual gland[1]. About 80% of parotid gland tumors are benign. The most common subtype of benign tumor is the pleomorphic adenoma, whereas the mucoepidermoid carcinoma is the most common malignant tumor[1]. Clinical examination, ultrasound scanning with or without fine needle aspiration cytology, and pre-operative CT-scan are the most commonly used methods to diagnose tumors of the salivary glands. Treatment of these tumors involves surgical resection[1].

In recent years, the study of salivary biomarkers has been taking flight and has evolved from the field of oral diseases[2, 3] to the field of systemic diseases[4-7]. The development of transcriptomic and proteomic approaches to study such diseases has provided the means to globally profile disease-associated molecules such as metabolites, proteins, DNA, mRNA and microRNA.

MicroRNA (miRNA) is a group of small non-coding RNAs, consisting of 19-25 nucleotides, which are involved in various cellular processes such as cell differentiation, proliferation and survival. They exert their effect by inhibiting translation through binding to complementary sequences in the 3' UTR of multiple target mRNAs, usually resulting in their silencing. MiRNAs are stably expressed in serum, plasma, urine, saliva, and other body fluids[8, 9]. Some miRNA species are specifically expressed in certain body fluids[9]. The unique expression patterns of miRNAs are correlated with certain human diseases, including various types of cancers. MiRNAs are frequently deregulated in cancer and have shown great promise as tissue-based markers for the classification of malignancies[10, 11].

Thus far, few papers have been published regarding the expression of miRNAs in salivary gland tumors tissue. It was reported that 22 miRNAs were expressed differently in pleomorphic adenomas when compared to the matched normal controls[12]. The majority of these differently expressed miRNAs were up-regulated in tumor tissue compared to the surrounding normal salivary gland tissue. Only a few miRNAs were down-regulated in tumor tissue. These down-regulated miRNAs were involved in the regulation of the pleomorphic adenoma gene. This is a proto-oncogene that is frequently overexpressed in various neoplasms, including pleomorphic adenoma. Another study described the expression of miRNA indirectly by studying the expression of Dicer in mucoepidermoid carcinoma[13]. Dicer is an Rnase III related

enzyme required for miRNA maturation that was differently expressed in mucoepidermoid carcinoma compared to the surrounding normal tissue. The abnormal expression of Dicer was significantly more frequent in high grade mucoepidermoid carcinoma correlating with lower survival rates[13]. From these studies, it can be concluded that the miRNAs machinery is deregulated, resulting in differences between miRNA profiles of salivary gland tumors and of normal tissue.

In this study, we investigated miRNA profiles in saliva from patients with malignant tumors and from patients with benign parotid gland tumors. Differences in miRNA profiles may be helpful to assist the clinical diagnosis of tumors in salivary glands. We report that significant different miRNA profiles were found between saliva from patients with a malignant and saliva from patients with a benign parotid gland tumor. The discovered salivary biomarkers possess inherent discriminatory potential for a non-invasive diagnostic tool for malignant and benign parotid gland tumors.

Methods and Materials

Patients

Whole saliva samples from patients with malignant (n=38) or benign parotid gland tumors (n=29) were obtained from the Salivary Gland Tumor Biorepository at the MD Anderson Cancer Clinic in Houston, TX. Samples were immediately stored at -80°C until ready for use. Before the samples were used, they were defrosted on ice and centrifuged for 10 min, 5000 rpm at 4°C. The cell-free supernatant was collected from the pellet and used immediately in the next step.

In the discovery phase**,** samples of whole saliva from 10 patients with malignant and samples of whole saliva from 10 patients with benign parotid gland tumor were matched (gender and ethnicity) and analyzed. Both benign and malignant groups consisted of: 7 men and 3 women. The mean age was 53 (33-82) and 60 (49-74) of the benign and the malignant tumor group, respectively. Both groups consisted of the same ethnic background (1 Hispanic, 1 Black, 8 Caucasians). An independent sample set was used for validation study (19 benign and 28 malignant). Smoking/drinking habits or clinicopathological data were neither recorded nor provided by the SGTB and were irretrievable. Patient's characteristics of the discovery and validation sample set are given in Table 1.

Table 1: Patients' characteristics

Discovery phase:			Validation phase:		
	Benign	**Malignant**		**Benign**	**Malignant**
Mean age (range)	53 (33-82)	60 (49-74)	Mean age (range)	57 (22-88)	57 (19-85)
sex (m/f):	6/4	6/4	sex (m/f):	14/5	19/9
Ethinicity			Ethinicity		
Hispanic	1	1	Hispanic	0	2
Caucasian	8	8	Caucasian	19	22
Black	1	1	Black	0	4
Tumor subtypes:			Tumor subtypes:		
Pleomorphic adenoma	6		Pleomorphic adenoma	13	
Warthin tumor	3		Warthin tumor	5	
Oncocytoma	1		Basal cell adenoma	1	
Neuroenocrine carcinoma		1	Squamous cell carcinoma		9
Oncocytic carcinoma		1	Adenocarcinoma NOS		2
Acinic cell carcinoma		1	Acinic cell carcinoma		3
Mucoepidermoid carcinoma		1	Mucoepidermoid carcinoma		3
Salivary duct carcinoma		2	Salivary duct carcinoma		3
Myoepthelial carcinoma		1	Undifferentiated carcinoma		1
Carcinoma ex-pleomorphic adenoma		1	Carcinoma ex-pleomorphic adenoma		3
Cystadenocarcinoma		1	Renal cell carcinoma		1
Adenoid cystic carcinoma		1	Adenoid cystic carcinoma		2
			Myoepithelial carcinoma		1
			Basal cell carcinoma		1

Salivary miRNA profiling and data analysis

Total RNA was isolated from 300 µl saliva supernatant using RNA extraction kits (Ambion mirVana Paris kit). DNase I treatment (DNase I, Qiagen) was used to remove contaminating DNA during RNA extraction. The concentration of total RNA was measured using RiboGreen assay. Extracted RNA (1-10 ng) was reverse transcribed and pre-amplified using the Taqman MicroRNA reverse Transcription Kit, Taqman PreAmp master mix, and Megaplex Primers (Applied Biosystems, Foster City CA). The pre-amplification product was not diluted prior to performing miRNA quantification. For the profiling of 750 miRNAs, a total volume of 105 µl was loaded into each well of the Taqman Human MicroRNA cards (Applied Biosystems,

Foster City CA), which were spun and run on the Applied Biosystems 7900HT Fast Real-Time PCR instrument containing a special card holder (Applied Biosystems, Foster City CA). Using default Taqman low density array setting and 6-carboxyfluorescein as a reporter, the RT-qPCR reaction was run at 95°C for 10 min to activate the enzyme and was then followed with 40 cycles at 95°C for 15 s and 60°C for 60 s.

The Ct value is defined as the cycle number in the fluorescence emission which exceeds that of a fixed threshold. A Ct of 15-30 was considered high expression and a Ct of 35 is considered low expression. A Ct value above 40 was considered as undetectable miRNA. For miRNA qPCR experiments, U6 snRNA was used as the reference gene. We calculated ΔCt by subtracting the Ct value of the reference gene (RNA polymerase III transcribed U6 snRNA) from the Ct value of each candidate biomarker. Data normalization was performed using RQ manager 1.2.1 and Data Assist v3.0 from Applied Biosystems. The qPCR based gene expression values between the two groups were compared using the non-parametric Wilcoxon test. Potential miRNA genes were then selected based on $P < 0.05$.

Verification of salivary miRNA markers

From the biomarker candidates generated by the Taqman MicroRNA array cards, 19 were verified by real-time quantitative PCR using Taqman MicroRNA assays on the same set of samples used in the discovery. Total RNA were reverse-transcribed with the Taqman MicroRNA Reverse Transcription kit using the following thermal cycling conditions: 16°C for 30 min, 42°C for the 30 min, 85°C for 5 min and then cooling to 4°C. Pre-amplification was performed with Taqman PreAmp master mix using the following thermal cycling conditions: 95°C for 10 min, 55°C for 2 min, 72°C for 2 min, 72°C for 2 min, 12 cycles at 95°C for 15 s and 60°C for 4 min, then 99.9°C for 10 min to inactivate the enzyme, and then ending at 4°C. qPCR was carried out in 384 well plate in reaction volume of 10 µl using Taqman Universal PCR Master mix with no UNG. Initial denaturing was performed for 10 min at 95°C and then followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min on the Roche LightCycler 480 II (Roche, San Francisco, CA). All RT-qPCRs were performed in duplicate for all candidate miRNA. Samples in which miRNA was omitted were used as negative controls. The Ct was examined and the ΔCt was calculated. For miRNA qPCR experiments, U6 snRNA was used as the reference gene.

Validation of salivary miRNA markers

Of the verified biomarker candidates, 9 were validated in an independent sample group of saliva samples from 19 patients with a benign parotid gland tumor and from 28 patients with a malignant parotid gland tumor. Total RNA was isolated as described above. Validation was done using the Taqman microRNA assays as described above for the verification of miRNAs. All qPCRs were performed in duplicate for all candidate miRNA and negative controls (in which RNA was omitted). The Ct was examined and the ΔCt was calculated. For miRNA qPCR experiments, U6 snRNA was used as the reference gene.

*Statistical methods_*

Wilcoxon rank-sum test was performed to compare the expression of miRNA in samples from patients with malignant versus benign parotid gland tumors in order to determine the level of significance for the miRNA biomarkers. Next we carried out a preclinical validation study on an independent set of 47 saliva samples from 19 patients with a benign and 28 with a malignant parotid gland tumor. The Wilcoxon rank-sum test was used to compare the ΔCt values for miRNAs between malignant and benign. Next, multivariate logistic regression analysis was used to construct a classification model to discriminate between patients with a malignant tumor and patients with a benign tumor. Forward stepwise model selection criterion was used to obtain a final model. ROC curves were constructed to determine the diagnostic/predictive values of individual as well as combined biomarkers from the logistic model. The performance of the model for classification was assessed by identifying the cut-off point of the predicted probability which yielded the largest sum of sensitivity and specificity.

Results

Discovery and verification of salivary transcriptomic markers

During biomarker discovery, expression of 750 miRNAs was screened using Taqman Human MicroRNA card. Of the 750 miRNA, 57 miRNAs were differently expressed in saliva from patients with a benign tumor (n=10) compared to patients with a malignant parotid gland tumor (n=10) (Wilcoxon test, $P < 0.05$). The expression of 54 out of the 57 miRNAs was higher in saliva samples from patients with a malignant tumor compared to expression in saliva samples from patients with a benign tumor. The expression of only 3 miRNAs was lower in saliva samples from patients with a malignant tumor compared to saliva samples from patients with a benign parotid gland tumor (hsa-miR-519b-3p, hsa-miR-520C-3p, hsa-miR-520D-3p). These 57 miRNA changes are unlikely due to chance alone ($P < 0.05$; Table S1 and S2).

RT-qPCR was performed on part of the discovery sample set to verify the Taqman Human MicroRNA card results. Eighteen miRNAs were chosen for verification in the original sample set. Selection of miRNAs was based on the *P* –value < 0.05 and fold change (Table S2). Of these eighteen miRNAs, 4 miRNAs (mmu-miR-140-5p, hsa-miR-374, hsa-miR-222, and hsa-miR-15b) differed statistically significant (p<.05) and 5 miRNAs (hsa-let-7g, hsa-miR-132, hsa-miR-519b-3p, hsa-miR-223 and hsa-miR-30a-3p) showed a trend (0.08 > P > 0.05) (Table 2).

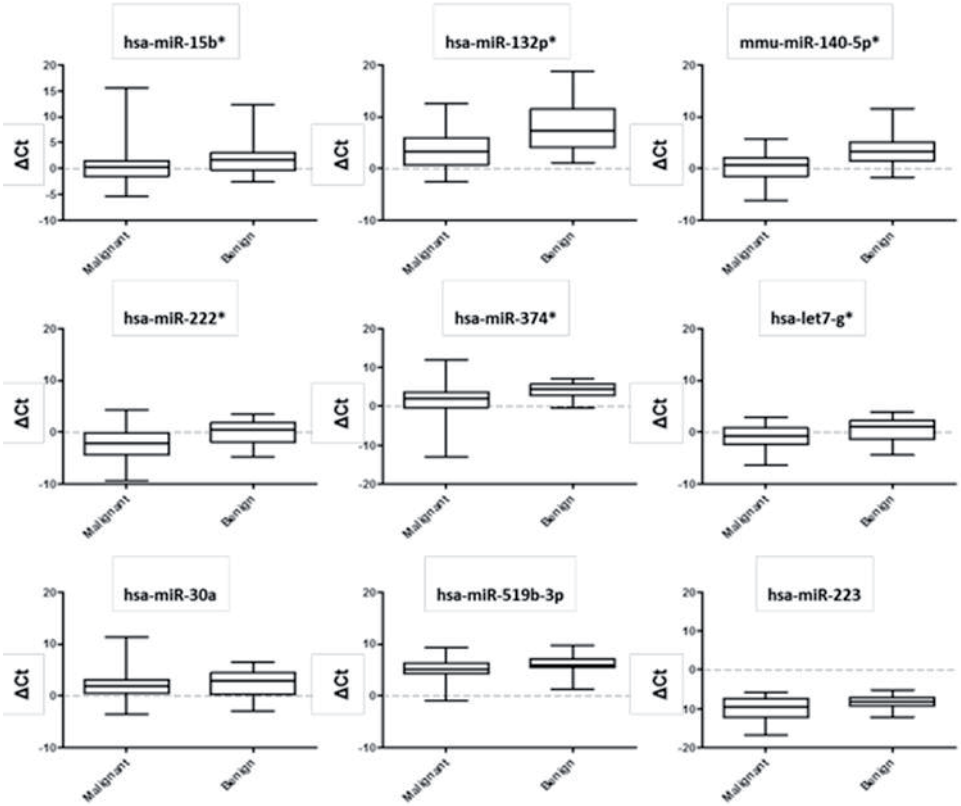
miRNA	Wilcoxon 2-sided malignant vs. benign
mmu-miR-140-5p	0.020**
hsa-miR-374	0.021**
hsa-miR-222	0.040**
hsa-miR-15b	0.046**
hsa-let-7g	0.050**
hsa-miR-132	0.052*
hsa-miR-519b-3p	0.054*
hsa-miR-30a-3p	0.070*
hsa-miR-223	0.079*
hsa-miR-20a	0.296
hsa-miR-195	0.384
hsa-miR-194	0.452
hsa-miR-618	0.534
dme-miR-7	0.6
hsa-miR-489	0.75
hsa-miR-509-5p	0.789
hsa-miR-381	0.808
hsa-miR-1285	0.913

Table 2: Verified salivary miRNA biomarkers in the original sample set. This set consisted of 10 whole saliva samples from patients with a benign parotid gland tumor and 10 whole saliva samples from patients with a malignant parotid gland tumor. * 0.08> P > 0.05; ** P < 0.05.

Validation of salivary miRNA markers

Next, the expression of 9 miRNAs (mmu-miR-140-5p, hsa-miR-374, hsa-miR-222, hsa-miR-15b, hsa-let-7g, hsa-miR-132, hsa-miR-519b-3p, hsa-miR-223 and hsa-miR-30a-3p) that were statistically significant or showed a trend was determined in a separate independent sample set consisting of saliva from 19 patients with a benign tumor and 28 patients with a malignant parotid gland tumor. The expression of all 9 validated miRNAs was increased in saliva samples from patients with a malignant parotid gland tumor, just as they were in the verification phase. Of the 9 miRNAs validated, the expression of 6 miRNAs (hsa-miR-374, hsa-miR-222, hsa-miR-15b, hsa-let-7g, hsa-miR-132, and mmu-miR-140-5p) was statistically significantly different between the two groups ($P < 0.05$) (Table 3; Figure 1; Table S3).

Figure 1. Box-and-whisker plot of validated miRNA expression profiles in saliva samples from patients with benign (n=19) and malignant (n=28) tumor in the parotid gland. Whiskers represent maximum and minimum ΔCt. * $P < 0.05$



miRNA	Wilcoxon 2-sided malignant vs. Benign
mmu-miR-140-5p	0.0003*
hsa-miR-374	0.0022*
hsa-miR-222	0.0070*
hsa-miR-15b	0.0019*
hsa-let-7g	0.0306*
hsa-miR-132	0.003*
hsa-miR-519b-3p	0.1931
hsa-miR-30a-3p	0.232
hsa-miR-223	0.0542

Table 3: Validated salivary miRNA biomarkers in an independent sample set. This set consisted of 19 whole saliva samples from patients with a benign parotid gland tumor and 28 whole saliva samples from patients with a malignant parotid gland tumor. * $P < 0.05$.

Evaluation of validated miRNA biomarkers

ROC curves were constructed to determine the diagnostic/predictive values of combined biomarkers from a logistic model. The performance of the model was assessed by identifying the cut-off point of the predicted probability which yielded the largest sum of sensitivity and specificity. A combination of hsa-miR-132, hsa-miR-15b, mmu-miR-140 and hsa-miR-223 yielded the highest values for sensitivity and specificity of 69% and 95%. The AUC for the 4-validated-miRNA combination was 0.90 (Figure 2). Of the validation data, a box-and-whisker plot predicting the probability for malignancy (Figure 3) was made based on the 4-validated miRNA combination (hsa-miR-132, hsa-miR-15b, mmu-miR-140, and hsa-miR-223). In the malignant tumor group, 14 of the 28 samples scored around the probability of malignancy of 0.95. This group consisted of different subtypes of malignant tumor (6 squamous cell carcinoma, 2 acinic cell carcinoma, 2 carcinoma ex-pleomorphic, 1 mucoepidermoid carcinoma, 1 adenoma not otherwise specified, 1 myoepithelial carcinoma, and 1 renal cell carcinoma). In the benign tumor group 8 from the 19 samples scored around the malignancy probability of 0.02. This group consisted of 8 samples (4 pleomorphic adenoma, 3 Warthin's tumor and 1 basal cell adenoma).

Figure 2. ROC curve computed from final logistic regression model. This model includes hsa-miR-132, hsa-miR-15b, mmu-miR-140, and hsa-miR-223, which in combination provided the best prediction. The area under the ROC curve was 0.9, the specificity was 95% and sensitivity was 69%.

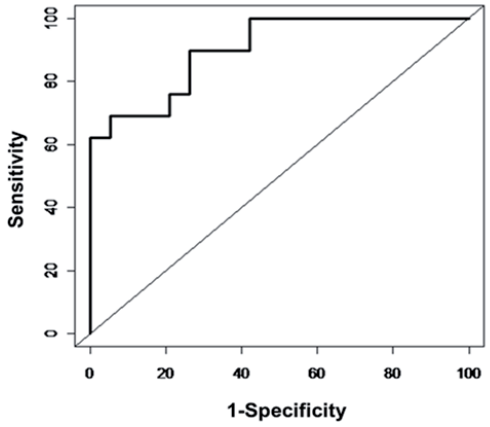
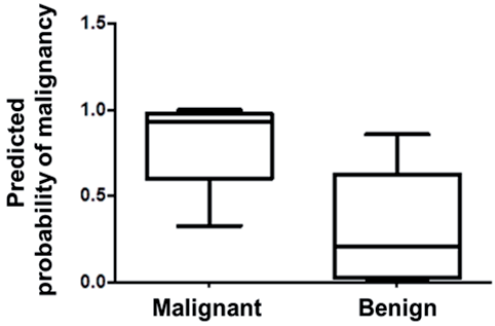


Figure 3. Box-and-whisker plot showing the "predicting probability for malignancy" based on the validation data of 4 miRNAs (hsa-miR-132, hsa-miR-15b, mmu-miR-140, and hsa-miR-223). Saliva samples from patients with a malignant tumor in the parotid gland have a higher "predicting probability for malignancy" than saliva samples from patients with a benign tumor in the parotid gland. Whiskers represent maximum and minimum probability for malignancy.



Discussion

Saliva has been recognized as an emerging diagnostic fluid[14-18]. It is readily available, and collection is relatively simple, inexpensive and, non-invasive. Using omics techniques a scientific basis has been laid underpinning its credential as a diagnostic fluid[3, 7, 10, 19-21]. A disadvantage of saliva as a diagnostic fluid is the notion that informative analytes are present in lower amounts in saliva than in serum[21]. Over the last few years, there has been increasing interest in the potential use of transcriptomic biomarker detection for cancer in bodily fluids[7, 19, 20, 22-24]. From this interest came the development of new and more sensitive technologies that can measure very low levels of analytes in saliva. Because of these new developments the low levels of analytes were no longer a drawback[25-27].

One of these analytes is miRNA. These small non-coding RNAs are stably expressed in all body fluids[9]. MiRNAs are also less prone to being degraded, unlike mRNA or proteins. This stable expression of miRNAs and the fact that miRNAs are less prone to be degraded make miRNAs a good choice for biomarkers[28, 29].

Cancer associated changes, such as chromosomal alterations, chromosomal losses or gains, mutation in miRNA gene and methylation of miRNA promoter often lead to changes in gene expression patterns. These consistent and validatable changes in miRNA gene expression can be used for diagnostic purposes[30]. The role of miRNAs in cancer can be two-sided. On one side, over-expression of miRNAs can result in the inhibition of tumor suppressor gene expression by binding to the target mRNA. This can lead to suppression of the tumor suppressor gene, which results in: increased proliferation, reduced apoptosis, increased invasiveness and angiogenesis. On the other side, miRNAs expression can be reduced. This can result in an over-expression of oncogenes and increased proliferation, reduced apoptosis, increased invasiveness and angiogenesis[30].

In this paper, we have identified and pre-validated the miRNAs (mmu-miR-140-5p, hsa-miR-374, hsa-miR-222, hsa-miR-15b, hsa-let-7g and, hsa-miR-132) that were differently expressed between saliva samples of patients with a malignant tumor and benign parotid gland tumor. Hsa-miR-15b has caspases 3, 8, 9 as validated targets. These proteins are involved in apoptosis pathways. Activation of caspase 3 by proteolytic cleavage due to activated caspase 8 and caspase 9 leads to irreversible commitment to apoposis [31]. Inhibition of these caspases by hsa-miR-15b will result in blockage of the apoptosis pathway. Another target of hsa-miR-15b is the reversion-inducing cysteine-rich protein with Kazal motifs (RECK) gene. RECK encodes for a glycosylphosphatidylinositol anchor glycoprotein and is an important inhibitor of matrix metalloproteinases. The expression of RECK is frequently reduced in carcinomas, e.g. colorectal[32, 33] and gastric cancer[34]. The reduced RECK expression is often correlated with poor prognosis.

Amongst the validated targets of let-7g are genes which are associated with cancer such as RAS, MYC, and CDKN2A. CDKN2A, also known as p16, is a tumor-suppressor-gene which is frequently deleted in a wide variety of cancers. Deletion of this gene can result in

increased cell proliferation[35]. RAS and MYC are oncogenes and are frequently mutated in cancer. These mutations in the RAS and MYC genes may have a negative effect on the binding of hsa-let-7g to its target mRNA, lowering the inhibiting effect of hsa-let-7g, which may lead to overexpression of RAS and MYC. This overexpression may result in an increased cell proliferation, demonstrating the double role miRNAs can play in cancer.

Fine-needle aspiration cytology (FNAC) is one of the techniques currently used to diagnose salivary gland neoplasms. A systematic review investigating the performance of FNAC in parotid gland lesions concluded that FNAC had a specificity of 97% and a sensitivity of 80%. However the performance variability was relatively large[36]. Therefore, additional (molecular) markers, such as those identified in the present study, can add to the accuracy of diagnoses on FNAC.

This study investigated and pre-validated the differences of miRNA expression in whole saliva from patients with malignant and benign parotid gland tumor. There appears to be a general up-regulation of miRNAs in saliva from patients with a malignant parotid gland tumor when compared to miRNA expression in saliva from patients with a benign parotid gland tumor. The AUC of the 4-validated-biomaker combination was 0.9 with a high specificity of 95% and a sensitivity of 69%.

There is some overlap in expression values; however, by using a combination of miRNA markers we improved the specificity and therefore the clinical usefulness of the diagnostic test. Furthermore, both the biomarker discovery and biomarker validation were performed using a variety of malignant parotid gland tumor types (18 subtypes). These clinical performances can be improved if we carry out the biomarker development using a specific subtype of salivary gland tumor (e.g. mucoepidermoid carcinoma).

The data that are presented in this study are encouraging toward developing a clinical application to distinguish malignant from benign parotid gland tumors. Because a large percentage of the tumors in the salivary gland arise in the parotid gland, investigating parotid saliva on the presence of these miRNA will be the next step. Also investigating differences in miRNA expression between whole saliva from patients with a salivary gland tumor and whole saliva from healthy people needs be a part of the additional exploration before the data's full clinical utility may be realized.

Disclosure of Potential Conflict of interest

David Wong is co-founder of RNAmeTRIX Inc., a molecular diagnostic company. PeriRx LLC sublicensed intellectual properties pertaining to molecular diagnostics from RNAmeTRIX. David Wong is a consultant to PeriRx.

References

1. Seifert G, Brocheriou C, Cardesa A, et al. WHO International Histological Classification of Tumours. Tentative Histological Classification of Salivary Gland Tumours. Pathology Research and Practice. 1990; 186:555-81.
2. Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009; 15:5473-7.
3. Wong DT. Salivary diagnostics. Journal of the California Dental Association. 2006; 34:283-5.
4. Gao K, Zhou H, Zhang L, et al. Systemic disease-induced salivary biomarker profiles in mouse models of melanoma and non-small cell lung cancer. PLoS One 2009; 4:e5875.
5. Parisi MR, Soldini L, Di Perri G, et al. Offer of rapid testing and alternative biological samples as practical tools to implement HIV screening programs. New Microbiology. 2009; 32:391-6.
6. Streckfus CF, Bigler LR, Zwick M. The use of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry to detect putative breast cancer markers in saliva: a feasibility study. Journal of Oral Pathology and Medicine. 2006; 35:292-300.
7. Zhang L, Farrell JJ, Zhou H, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology. 2010; 138:949-57.
8. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences of the U.S.A. 2008; 105:10513-8.
9. Weber JA, Baxter DH, Zhang S, et al. The microRNA spectrum in 12 body fluids. Clinical Chemistry. 2010; 56:1733-41.
10. Brinkmann O, Wong DT. Salivary transcriptome biomarkers in oral squamous cell cancer detection. Advances in Clinical Chemistry. 2011; 55:21-34.
11. Ueda T, Volinia S, Okumura H, et al. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncology. 2010; 11:136-46.
12. Zhang X, Cairns M, Rose B, et al. Alterations in miRNA processing and expression in pleomorphic adenomas of the salivary gland. International Journal of Cancer. 2009; 124:2855-63.
13. Chiosea S, Barner E, Lai S, et al. Mucoepidermoid carcinoma of upper aerodigestive tract: clinicopathologic study of 78 cases with immunohistochemical analysis of Dicer expression. Virchow Archives. 2008; 452:629-35.
14. Lee JM, Garon E, Wong DT. Salivary diagnostics. Orthodontics & Craniofacial Research. 2009; 12:206-11.
15. Lee YH, Wong DT. Saliva: an emerging biofluid for early detection of diseases. American Journal of Dentistry. 2009; 22:241-8.
16. Liu J, Duan Y. Saliva: A potential media for disease diagnostics and monitoring. Oral Oncology. 2012; 48:569-77.
17. Malamud D. Saliva as a diagnostic fluid. Dental Clinics of North America. 2011; 55:159-78.
18. Miller CS, Foley JD, Bailey AL, et al. Current developments in salivary diagnostics. Biomarkers in Medicine. 2010; 4:171-89.

19. Li Y, St John MA, Zhou X, et al. Salivary transcriptome diagnostics for oral cancer detection. Clinical Cancer Research. 2004; 10:8442-50.
20. Zhang L, Xiao H, Karlan S, et al. Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. PLoS One. 2010; 5:e15573.
21. Miller SM. Saliva testing--a nontraditional diagnostic tool. Clin Lab Sci 1994; 7:39-44.
22. Hu S, Arellano M, Boontheung P, et al. Salivary proteomics for oral cancer biomarker discovery. Clinical Cancer Research. 2008; 14:6246-52.
23. Lee EJ, Gusev Y, Jiang J, et al. Expression profiling identifies microRNA signature in pancreatic cancer. International Journal of Cancer. 2007; 120:1046-54.
24. Li Y, Elashoff D, Oh M, et al. Serum circulating human mRNA profiling and its utility for oral cancer detection. Journal of Clinical Oncology. 2006; 24:1754-60.
25. Juusola J, Ballantyne J. Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. Forensic Science International. 2003; 135:85-96.
26. Juusola J, Ballantyne J. mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. Journal of Forensic Science. 2007; 52:1252-62.
27. Li Y, Zhou X, St John MA, et al. RNA profiling of cell-free saliva using microarray technology. Journal of Dental Research. 2004; 83:199-203.
28. Streckfus CF, Dubinsky WP. Proteomic analysis of saliva for cancer diagnosis. Expert Revie of Proteomics. 2007; 4:329-32.
29. Thomadaki K, Helmerhorst EJ, Tian N, et al. Whole-saliva proteolysis and its impact on salivary diagnostics. Journal of Dental Research. 2011; 90:1325-30.
30. Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. Journal of Clinical Oncology. 2009; 27:5848-56.
31. Fan TJ, Han LH, Cong RS, et al. Caspase family proteases and apoptosis Acta Biochimica et Biophysica Sinica (Shanghai). 2005; 37:719-27.
32. Span M, Moerkerk PT, De Goeij AF, et al. A detailed analysis of K-ras point mutations in relation to tumor progression and survival in colorectal cancer patients. International Journal of Cancer. 1996; 69:241-5.
33. Takeuchi T, Hisanaga M, Nagao M, et al: The membrane-anchored matrix metalloproteinase (MMP) regulator RECK in combination with MMP-9 serves as an informative prognostic indicator for colorectal cancer. Clinical Cancer Research. 2004; 10:5572-9.
34. Song SY, Son HJ, Nam E, et al. Expression of reversion-inducing-cysteine-rich protein with Kazal motifs (RECK) as a prognostic indicator in gastric cancer. European Journal of Cancer. 2006; 42:101-8.
35. Caldas C, Hahn SA, da Costa LT, et al. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. Nature Genetics. 1994; 8:27-32.
36. Schmidt RL, Hall BJ, Wilson AR, et al. A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions. American Journal of Clinical Pathology. 2011; 136:45-59.

6.

The discovery of predictive salivary miRNAs as biomarkers for salivary gland tumors via statistical machine learning.

Johannes H. Matse, Evgeni Levin, Sultan Imangaliyev, Jan G.M. Bolscher, Roy Montijn, Bart Keijser, David T.W. Wong, Elisabeth Bloemena and Enno C.I. Veerman

Submitted for publication in *PLoS One*

Abstract

Prospective-specimen-collection-retrospective-blinded-evaluation design is a rigorous methodology for biomarker discovery, but cannot always be implemented practically. Validating biomarkers in an independent sample set may be problematic when the disease is rare and the amount of samples is limited (e.g. salivary gland tumors). Therefore, we use an alternative design to identify and statistically validate predictive salivary micro-RNAs in whole saliva from 20 patients with a salivary gland tumor. Using machine-learning-techniques, we were able to construct a signature of seven salivary miRNAs that distinguished benign from malignant parotid gland tumors (specificity 86%; sensitivity 91%; AUC 0.92). Hsa-miR-324-5p, hsa-miR-374 and hsa-miR-1285 were increased in saliva from patients with malignant tumors, while hsa-miR-411, hsa-miR-449a, hsa-miR-449b and hsa-miR-599 were increased in saliva from patients with benign tumors . Hsa-miR-449a and hsa-miR-449b repress cell-cycle, while hsa-miR-374 and hsa-miR-1285 accelerate cell-cycle. The proposed approach is particularly suitable for biomarker-selection and discovery. Additionally these findings may lead to explore the role of these miRNAs in respect to in the cell-cycle of salivary gland tumors.

Introduction

Salivary gland tumors are a rare but very heterogenic group of tumors comprised of 13 benign and 24 malignant subtypes. Approximately 80% of all salivary gland tumors occur in the parotid gland (1). Clinical examination, ultrasound scanning (with or without fine needle aspiration cytology), preoperative CT-scan and MRI are the most commonly used diagnostic methods for diagnosing these rare tumors[1]. Therefore, a non-invasive salivary diagnostic tool may be helpful in assisting the clinician in diagnosing salivary gland tumors.

MiRNA expression is often dysregulated in tumors which can have huge impact on the progress of the disease[2,3]. Therefore, in our previous study we profiled the miRNA expression in saliva from patients with a malignant or benign salivary gland tumor[4]. After validation 6 miRNAs (hsa-miR-15b, hsa-miR-222, hsa-miR-132, hsa-miR-374, hsa-mmu-140-5p and miR-let-9) were found to be statistically significantly different[4]. A four miRNA signature (hsa-miR-15b, hsa-miR-132, mmu-miR-140 and hsa-miR-223) was able to distinguish tumor type in whole saliva samples from patients with a malignant or a benign parotid gland tumor (specificity 95% and sensitivity 69%).

In the previous study, the research design was based on the prospective-specimen-collection, retrospective-blinded-evaluation (PRoBE) design which consists of discovery, verification and validation phases[5]. According to the PRoBE design, the discovered miRNAs are validated in an independent sample set[5]. Although such design is based on a rigorous methodology and valid biological assumptions, it cannot always be implemented practically. Validating biomarkers in an independent sample set may be problematic when the disease is very rare and the amount of samples is limited. Therefore, an alternative to biological validation should be used in cases of rare and very diverse diseases such as salivary gland tumors. This alternative approach should use all the available data in the most efficient way and replace expensive and long biological validation experiments with cheaper and faster techniques. The use of new statistical machine learning methods may be one of the possible replacements or additional tools to existing methodologies for rare diseases.

In recent years, new statistical machine learning methods have been used to analyze the growing amount of data that is available[6,7]. Due to outstanding performances in various data analysis tasks[8], they are particularly well suited for modeling complex biological datasets which are high-dimensional, noisy, relationally structured and heterogeneous[9]. As an alternative to PRoBE design-like approaches, machine learning techniques can statistically validate the biomarker signature that is revealed in the discovery phase by splitting the samples in a training and test set. Furthermore, unlike standard univariate statistical methods, statistical machine learning algorithms are able to capture complex group-based correlations in data, which are difficult to detect using standard statistical testing such as two-sided Wilcoxon test[10,11].

In this study, we used a combination of three statistical machine learning techniques to analyze the miRNA expression signatures of saliva samples from patients with a benign or malignant parotid gland tumor. These are Elastic net, stability selection and randomization

test. These three methods combined provide a rigorous testing of results and replaces in vitro biological validation with in silico validation based on the limited amount of data. Application of machine learning conducted on the collected data revealed a signature of the seven most robust miRNA. This preliminary group-based salivary miRNA signature could distinguish between benign and a malignant parotid gland tumor (specificity 86%; sensitivity 91%; AUC 0.92). Interestingly, four miRNAs in the signature have targets that are involved in the regulation of the cell cycle . This approach seems particularly suitable for detection of miRNAs expression signatures in highly-dimensional data sets of high biological diversity such as salivary gland tumors. In addition, machine learning techniques are of value in cases where the number of samples available for the analysis is limited due to the rarity of the disease.

Methods and Materials

MiRNA data set

The miRNA expression data set was created by profiling miRNAs by real-time PCR using the TaqMan Human MicroRNA Cards (Applied Biosystems, Waltham, Massachusetts, USA) containing 750 miRNA (Matse et al, 2013). For profiling of miRNAs 20 whole saliva samples were used from patients with malignant (N =10) or benign parotid gland tumors (N =10). The saliva samples used, were obtained from the Salivary Gland Tumor Biorepository at the MD Anderson Cancer Clinic in Houston, TX (Table 1).

Table 1: Patients' characteristics of patients with a benign (N =10) or malignant (N =10) parotid gland tumor.

		Benign	**Malignant**
Mean age (range)		53 (33-82)	60 (49-74)
sex (male/female)		6/4	6/4
Ethnicity			
	Hispanic	1	1
	Caucasian	8	8
	Black	1	1
Neoplasm subtypes			
	Pleomorphic adenoma	6	
	Warthins tumor	3	
	Oncocytoma	1	
	Neuroendocine carcinoma		1
	Oncocytic carcinoma		1
	Acinic cell carcinoma		1
	Mucoepidermoid carcinoma		1
	Salivary duct carcinoma		2
	Myoepithelial carcinoma		1
	Carcinoma ex-pleomorphic adnoma		1
	Cystadenocarcinoma		1
	Adenoid cystic carcinoma		1

Feature selection: a brief overview

Machine learning techniques including but not limited to feature selection have many examples of successful applications in molecular biology (7,8), because biological datasets are large, noisy and have much more features than examples. In biological applications,

feature selection techniques focus on finding the most important biomarkers in the data with respect to a certain disease or the pathological state of a biological system. It automatically selects the most relevant features e.g. miRNA, mRNA, DNA etc., while excluding the irrelevant ones without transforming the original values like PCA, so that the model interpretation becomes possible afterwards[12]. Feature selection is successful mainly due to three reasons[13]. First of all, it improves the prediction performance of the models due to removal of noisy irrelevant features.

Secondly, it provides a more compact list of predictors which improves speed and reduces the costs of data collection. Finally, it provides a better understanding of the underlying biological process reflected by generated data. The feature selection method used in this study combines elastic net and stability selection procedures. The results of the model were additionally checked by a randomization test.

Elastic Net

Elastic net simultaneously performs feature selection and classification, i.e. model learns from data a pattern which correctly discriminates benign tumor from malignant one while at the same time selecting only those minimum number of miRNAs which are the most relevant to the model outcome. Such an embedded feature selection approach also solves the so called 'p<<N' problem, in which the number of examples p (e.g. tumor samples) is much less than the number of features N (e.g. miRNAs). Other classical statistical methods are often numerically unstable and may give misleading results.

Success of Elastic net is possible due to its combination of the advantages of two regularization-based feature selection methods such as LASSO and Ridge regression[14]. Ridge regression has the tendency to select correlated features, while LASSO on the other hand reduces the number of features in the model by shrinking corresponding weighting coefficients to zero. More detailed information about elastic net is provided in supplementary material.

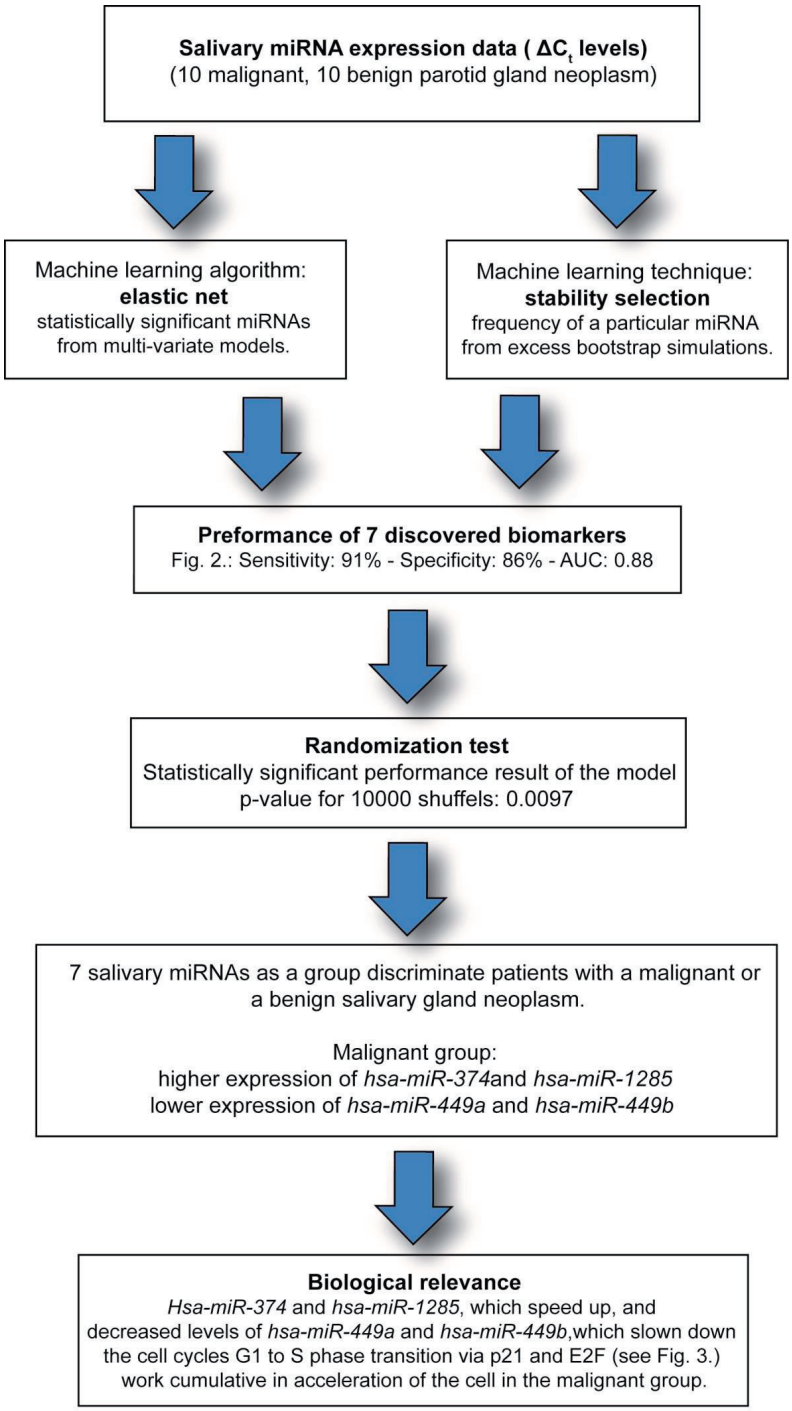
Stability selection

Stability selection[15] is a technique designed to improve the performance of a feature selection algorithm based on the sub-sampling of examples. The general idea of stability selection is to select only those miRNAs that are consistently selected throughout multiple runs of a feature selection model. In this study this is done by running elastic net many times on randomly subsampled data and choosing those variables that are selected most frequently across all sample partitions. More detailed information about stability selection procedure is provided in supplementary material.

Randomization test

After feature selection, we also performed an additional statistical test of the model results by using the randomization test procedure. We assumed that there is a connection between

Figure 1: Overview of Probe and machine learning research designs. Where the PRoBE design has discovery verification and validation phases with new sample sets, machine learning can statistically validate the markers without the use of a new sample set.



miRNA expression level values and corresponding tumor classification label values, i.e. benign or malignant. We tested this hypothesis by conducting modeling on the dataset with reshuffled class labels (benign or malignant) while keeping the corresponding miRNA expression value (delta Ct values) indices fixed. The null-hypothesis is that the order of labels for the classification test does not matter, meaning that the model will give a good AUC score even if the order of labels is shuffled, i.e. connection between miRNA values and corresponding labels is completely random and model simply learns noise having low predictive power on new previously unseen data. The alternative hypothesis is that the order of the labels for the classification test does matter, and that only the most predictive miRNA values' combination gives a good AUC score when correctly labeled and model indeed learns the correct pattern providing biologically relevant results. Such a procedure is often used to test the statistical validity of the results and is known as a randomization test[16]. Randomization test results suggested that we should reject null-hypothesis and accept the alternative one, meaning that feature selection results must be biologically relevant. More detailed information about randomization test is provided in supplementary material.

ROC AUC score

The performance measure used in this study to evaluate outcome of the model for a binary classification of tumor types is a Receiver Operating Characteristics Area Under Curve (ROC AUC)[17]. The ROC can be understood as a plot of the probability of correctly classifying the positive samples (benign group) against the rate of incorrectly classifying true negative samples (malignant group). The area under ROC curve value varies between 0.5 (completely random prediction) and 1.0 (prefect predictor) Thus, the numerical value of ROC AUC score is a convenient and balanced measure of the model's predictive accuracy.

Results

The miRNA data obtained in the discovery phase were used as input for the statistical learning algorithm. Specifically, we used the expression levels of 750 miRNA species in saliva samples from 20 patients (10 malignant and 10 benign).

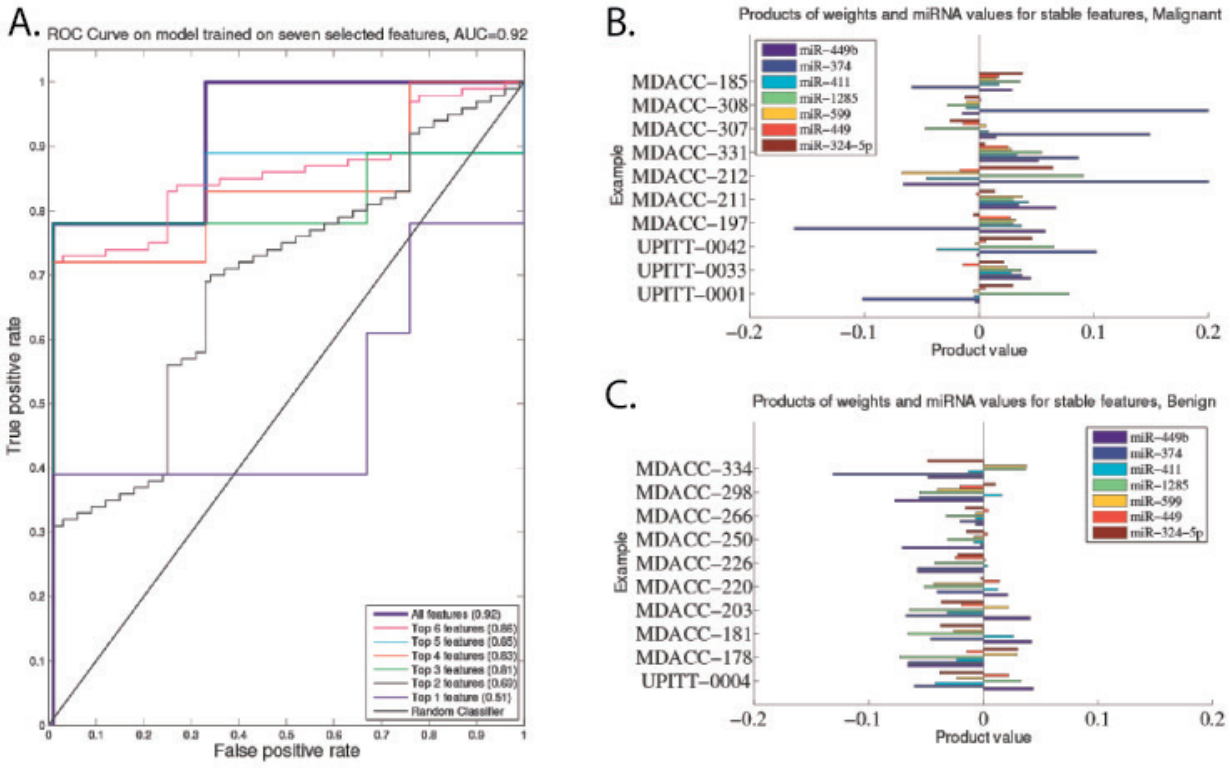
Stability selection procedure conducted on the miRNA expression profiles allowed us to select the most robust miRNA biomarkers. These biomarkers, as well as their stability coefficients, are presented in Table 2.

Table 2: Stability coefficients and average ΔCt for the discovered miRNAs.

miRNA	Stability coefficient	Average ΔCt Benign	Average ΔCt Malignant	Expression level
hsa-miR-449b	0.88	15.1	18.5	B>M
hsa-miR-374	0.65	7.2	0	M>B
hsa-miR-411	0.63	19.5	18.5	M>B
hsa-miR-599	0.57	16.3	21	B>M
hsa-miR-1285	0.56	16.4	9.6	M>B
hsa-miR-324-5p	0.46	16.5	11.8	M>B
hsa-miR-449a	0.44	14	15.9	B>M

Of the 750 analyzed miRNAs, hsa-miR-449b had the highest stability coefficient (0.88). Other miRNAs with high stability coefficients were hsa-miR-374, hsa-miR-411, hsa-miR-599, hsa-miR-1285, hsa-miR-324-5p and hsa-miR-449a. ROC curves were constructed to determine the diagnostic/predictive values of the combined 7 selected miRNAs, resulting in an AUC of 0.92 (specificity 86%; sensitivity 91%) (Figure 2A). This result was statistically significant ($P < 0.01$) by application of the randomization test (Supplementary Figure S2), demonstrating that the most stable biomarkers are also highly predictive in terms of the classification of the saliva samples into benign or malignant salivary gland tumor categories.

Figure 2: ROC curve and Bar chart of selected biomarkers' and weighting coefficients' products for malignant and benign tumor groups. A. Receiver-operating-characteristics (ROC) curve. ROC computed from the final model. The best prediction model included 7 miRNAs, which in combination provided the best prediction. The area-under-the-curve value of the ROC curve was 0.92, the specificity was 86% and sensitivity was 91%. B. The product values for the malignant parotid gland samples. C. The product values for the benign parotid gland samples. The weights for the malignant samples go in the opposite direction in comparison to the weights of the benign samples. The width of the bar corresponds to the value of the product of weights and delta-Ct.



Besides delta-Ct values (Supplementary Figure S1), estimated elastic net coefficients also play key a role in classification of tumor samples. To visualize the joint influence of the elastic net coefficients and delta-Ct, we plotted the product of the elastic net coefficient and delta-Ct for both the malignant and benign tumors samples (Figure 2B and C). The overall sum of all products per each sample resulted in a different pattern for each category (malignant vs benign). While, the overall sum of all products per sample for malignant tumors moved towards the positive direction, the opposite was true for the benign tumor group in which the overall sum of all products per each sample moved towards the negative direction, splitting the malignant from the benign salivary gland tumors.

Discussion

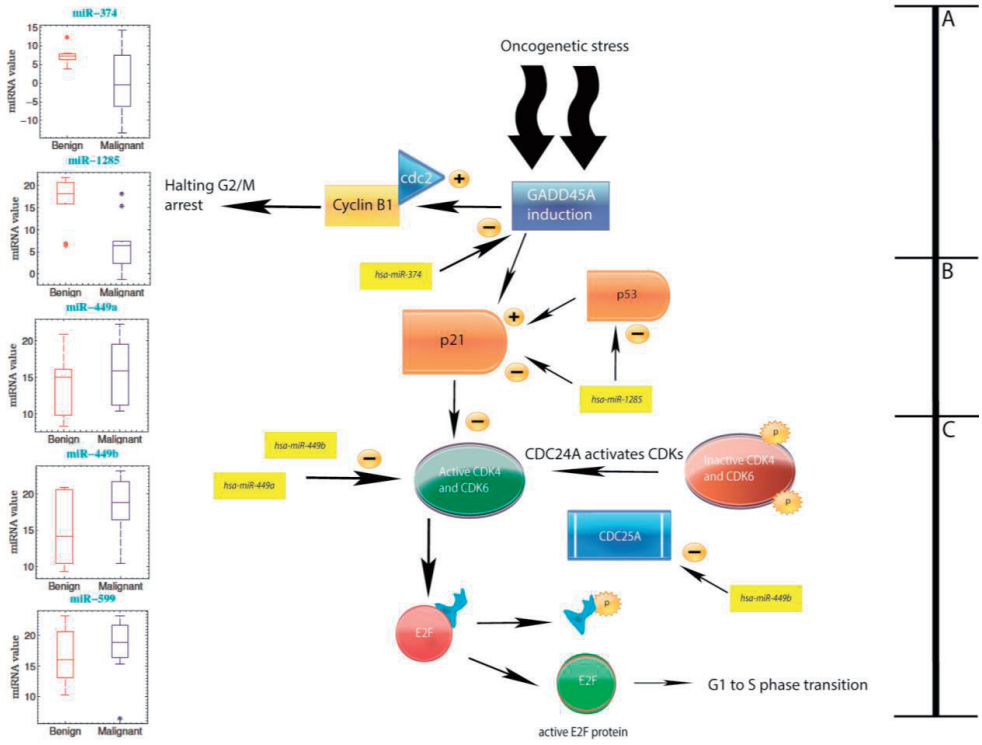
In this study, we applied novel statistical methods such as machine learning techniques for the discovery of prognostic miRNA biomarkers. Unlike PRoBE design, this methodology does not require using biological validation on external dataset, thus it is preferable for the diagnosis of rare diseases such as salivary gland tumors. With the use of machine learning techniques, such as elastic net, a new list of salivary miRNAs were discovered that could discriminate between patients with a malignant or benign salivary gland tumor (sensitivity 0.91; specificity 0.86; AUC 0.92).

In a former study[4] a combination of 4 salivary miRNAs was able to discriminate malignant from benign parotid gland tumors (sensitivity 0.69; specificity 0.95; AUC 0.90). That study was based on an assumption that miRNAs do not have any interaction with each other, therefore it is valid to use an univariate statistical test to discriminate tumor types. Using machine learning techniques in this study, we initially found an overlap with miRNAs that were discovered using the PRoBE design. The top 7 miRNAs, however, were different and were able to discriminate saliva from patients with a malignant or benign parotid gland tumor. This combination has a higher sensitivity compared to the miRNA combination found by the PRoBE design. The explanation for this may be found in the fact that in the PRoBE design uses filters to handle the amount of data. This leaves a part of the data untouched. Machine learning techniques, on the other hand, use all data points and therefore exploit the full potential of the data.

Furthermore, unlike the miRNAs discovered using the PRoBE design, some of the discovered miRNAs in this study seems to have a group based correlation. Four identified miRNAs (hsa-miR-374, hsa-miR-1285, hsa-miR-449a and hsa-miR-449b) have been implicated in the control of the cell cycle as a “gas- or brake-pedal” and their actions been validated in earlier studies[18-24] (Figure 3).

Hsa-miR-374 and hsa-miR-1285 act as “gas pedals” that speed up the cell cycle: hsa-miR-374 suppresses the expression of GADD45A, an inducer of cell cycle arrest[18,22]. Furthermore, both hsa-miR-374 and hsa-miR-1285 suppress p21[19,21]. This results in the suppression of the p21-mediated inhibition of CDK6 which in conjunction with CDK4 acts as a switch directing the cell towards the S-phase. It can be envisaged that increased levels of these miRNAs, e.g. in malignant tumors, will enhance the rate of cell-division.

Figure 3: Biological interpretations of the validated targets. Hsa-miR-374 inhibits the transcription of growth arrest and DNA-damage-inducible protein (GADD45A) (Section A)[18,19]. Hsa-miR-1285 inhibits the transcription of p21 directly by binding p21 mRNA, but also indirectly by inhibiting the transcription of p21 activator, p53 (Section B)[20, 21]. Hsa-miR-449a and hsa-miR-449b both can inhibit the transcription of CDK4 and CDK6 (Section C). Furthermore, hsa-miR-449b can inhibit the transcription of CDK4 and CDK6 activator, cell division cycle 25A (CDC25A)[22, 23]. A higher expression of hsa-miR-374 and hsa-miR-1285 and a lower expression of hsa-miR-449a and hsa-miR-449b in saliva from patients with a malignant tumor may result in an uncontrolled acceleration of the cell cycle in malignant tissues. The expression levels of the miRNAs in saliva from patients with a malignant and benign salivary gland tumors are portrait on the left hand side as box and whisker plots. Whiskers represent maximum and minimum delta-Ct., – inhibition; + activation.



On the other hand, hsa-miR-449a and hsa-miR-449b act as "brake-pedals" for cell-division by directly and indirectly suppressing CDK6 and CDK4[19,22,23]. We hypothesize that the increased levels of hsa-miR-374 and hsa-miR-1285 (the "gas-pedals") in combination with the decreased levels of hsa-miR-449a and hsa-miR-449b (the "brake-pedals") will result in an uncontrolled acceleration of the cell cycle in malignant tissues. Strikingly, without a priori mechanistic information, the statistical learning algorithm, using 750 miRNA-data as input, selected these four miRNAs that coordinately target the cell division cycle.

The three remaining miRNAs do not have any validated targets yet, but do have predicted targets. Hsa-mir-411 is predicted to have the pro-apoptotic BCL2 interacting protein 3

(BNIP3) as its target gene, which is often reduced in tumors. Furthermore, hsa-miR-599 seems to target the v-myc avian myelocytomatosis viral oncogene homolog or MYC gene[24] and acts as a tumor suppressor and inhibits hepatocellular carcinoma cells proliferation, migration and invasion by partly targeting oncogenic MYC[24].

Key aspects of our computational analysis are the application of the novel statistical learning methods, such as sparse regularization, and the stability selection procedures. This combination allows us to utilize the full potential of the gathered data. Specifically, the statistical learning algorithm builds multivariate models which can identify an ensemble of biomarkers characteristic of malignant tumors.

Unlike the PRoBE design, no new independent sample set is needed for the validation of the discovered miRNA using machine learning techniques. By splitting the sample set into a learning and a test sample set, the miRNAs can be statistically validated without needing new samples. This is extremely useful for rare diseases or for diseases in which samples are difficult to collect.

In this paper, we applied an appropriate and efficient machine learning technique for the discovery of salivary miRNAs that can discriminate saliva from patients with a benign and malignant parotid gland tumor. Even though additional validation of the miRNAs and their targets in salivary gland tumors is recommended based on methodologies such as PRoBE, these findings can serve as a basis upon which further research into the mechanistic aspects of salivary gland tumor pathology may be based. Furthermore, these biomarkers can also help to develop new clinical applications for a possible treatment of salivary gland tumors.

References

1. Seifert G, Brocheriou C, Cardesa A, et al. WHO International Histological Classification of Tumours. Tentative Histological Classification of Salivary Gland Tumours. Pathology Research and Practice. 1990; 186:555-81.
2. Chen PS, Su JL, Hung MC. Dysregulation of MicroRNAs in cancer. Journal of Biomedical Sciences. 2012; 19:90.
3. Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. Journal of Clinical Oncology. 2009; 27:5848-56.
4. Matse JH, Yoshizawa J, Wang X, et al. Discovery and pre-validation of salivary extracellular microRNA biomarkers panel for the non-invasive detection of benign and malignant parotid gland tumors. Clinical Cancer Research. 2013; 19:3032-8.
5. Pepe MS, Feng Z, Janes H, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. Journal of the National Cancer Institute. 2008; 100:1432-1438.
6. Mapstone M, Cheema AK, Fiandaca MS, et al. Plasma phospholipids identify antecedent memory impairment in older adults. Nature Medicine. 2014; 20:415-418.
7. Biesbroek G, Tsivtsivadze E, Sanders EA, et al. Early Respiratory Microbiota Composition Determines Bacterial Succession Patterns and Respiratory Health in Children. American Journal of Respiratory and Critical Care Medicine. 2014; 190:1283-92.
8. Cornelisse LN, Tsivtsivadze E, Meijer M, et al. Molecular machines in the synapse: Overlapping protein sets control distinct steps in neurosecretion. PloS Computation Biology. 2012; 8:e1002450.
9. Koller D, Friedman N. 2009. Probabilistic Graphical Models: Principles and Techniques. The MIT Press: Cambridge, Massachusetts, USA.
10. Imangaliyev S, Keijser B, Crielaard W, et al. In: Online Semi-supervised Learning: Algorithm and Application in Metagenomics. Paper presented at IEEE International Conference on Bioinformatics and Biomedicine (China), 2013.
11. Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning: data mining, inference and prediction. 2nd edition. Springer: New York, New York, USA.
12. Shalev-Shwartz S, Ben-David S. 2014. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press: New York, NY, USA.
13. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research. 2003; 3:1157-1182.
14. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B. 2005; 67:301-320.
15. Meinshausen N, Bühlmann P. Stability selection. Journal of the Royal Statistical Society: Series B. 2010; 72:417-473.
16. Fisher RA. 1935. The design of experiments. Oliver and Boyd: Edinburgh, UK.
17. Fawcett T. 2006. An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27:861-874.

18. Bueno MJ, Malumbres M. MicroRNAs and the cell cycle. Biochimica et Biophysica Acta. 2011; 1812:592-601.
19. Fan W, Richter G, Cereseto A, et al. Cytokine response gene 6 induces p21 and regulates both cell growth and arrest. Oncogene. 1999; 18:6573-6582.
20. Lizé M, Pilarski S, Dobbelstein M. E2F1-inducible microRNA 449a/b suppresses cell proliferation and promotes apoptosis. Cell Death & Differentiation. 2010; 17:452-458.
21. Tian S, Huang S, Wu S, et al. 2010. MicroRNA-1285 inhibits the expression of p53 by directly targeting its 3' untranslated region. Biochemical and Biophysical Research Communications. 2010; 396:435-439.
22. Wang XW, Zhan Q, Coursen JD, et al. GADD45 induction of a G2/M cell cycle checkpoint. Proceedings of the National Academy of Sciences USA. 1999; 96:3706-3711.
23. Yang X, Feng M, Jiang X, et al. miR-449a and miR-449b are direct transcriptional targets of E2F1 and negatively regulate pRb-E2F1 activity through a feedback loop by targeting CDK6 and CDC25A. Genes and Development. 2009; 23:2388-2393.
24. Tian J, Hu X, Gao W, et al. Identification a novel tumor-suppressive hsa-miR-599 regulates cells proliferation, migration and invasion by targeting oncogenic MYC in hepatocellular carcinoma. American Journal of Translational Research. 2016; 8:2575-2584.

Supplementary

Figure S1: Pie-chart showing the "Sum of ΔCt of the selected 7 miRNA". A. Pie-chart representing the distribution of the sum of ΔCt of the discovered miRNAs for malignant parotid gland tumors. B. Pie-chart representing the distribution of the sum of ΔCt of the discovered miRNAs for benign parotid gland tumors.

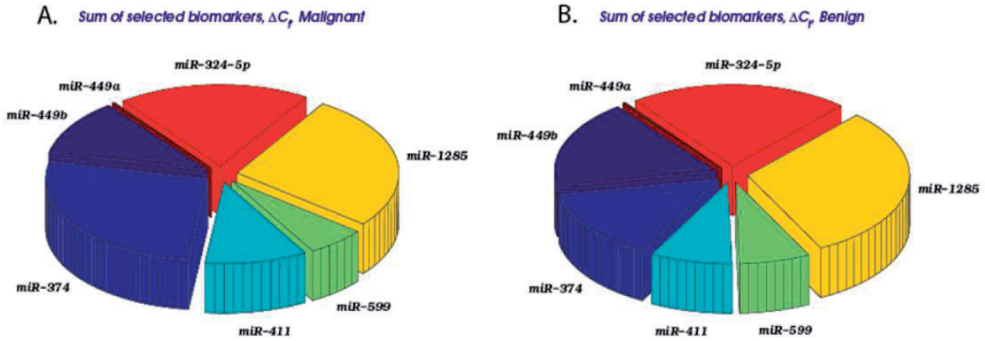
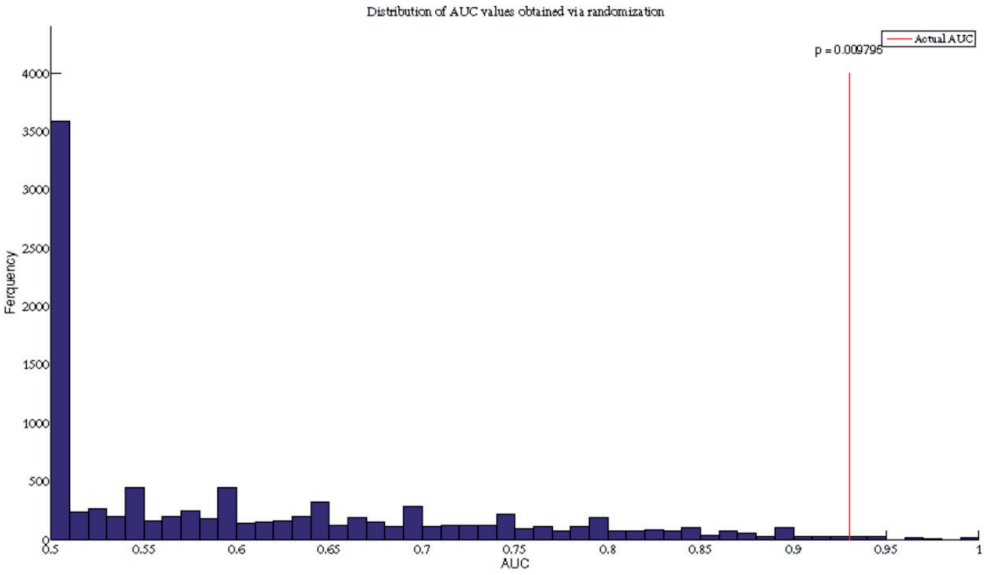


Figure S2: Histogram of AUC scores obtained via the randomization test after 10000 shuffles.



7.

General summary and future aspects

Salivary gland tumors are a heterogeneous class of tumors which despite clinical and technical advancements are difficult to diagnose. One major causes of this is the overlap in histopathology of these tumors. Other important causes are the rarity of salivary gland tumors, and the lack of molecular diagnostic markers that may help in identifying the tumors subtype. The global annual incidence of all salivary gland tumors (malignant and benign combined) ranges from 0.4-13.5 cases per 100,000 population worldwide and 3 cases per 100,000 population in the Netherlands.

Currently, clinical examination, ultrasound scanning with or without fine needle aspiration cytology, preoperative CT-scan and MRI are available for the differential diagnosis in parotid gland swelling[1]. The performance of fine needle aspiration cytology in parotid gland lesions has a specificity of 97% and a sensitivity of 80%. However, the performance variability is relatively large[2], so improving the accuracy, e.g. by utilization of additional (molecular, non-invasive) biomarkers, is still needed. Therefore, using novel biomarkers as an additional tool can add to the accuracy of diagnoses of salivary gland tumors.

This thesis aimed to find novel, molecular, non-invasive biomarkers for the profiling and diagnosis of salivary gland tumors. Biomarkers can be found at DNA, RNA and protein levels, and all three levels were addressed: (1) on the protein level through investigating the expression of mucins and mucin-associated carbohydrates by immunohistochemistry, (2) on the DNA level by cataloging genomic aberrations using arrayCGH, (3) on the RNA level by determining the miRNA expression levels in whole and parotid saliva by RT-qPCR.

Altered expression of mucins and mucin–type carbohydrates in mucoepidermoid carcinoma.

Mucoepidermoid carcinoma belong to the most common malignant salivary gland neoplasms, consisting of mucous, intermediate and epidermoid cells and are classified based on histological characteristics as low, intermediate or high grade tumors. Chapter 2 demonstrates the abnormal expression and localization of mucins and mucin-type carbohydrates in mucoepidermoid carcinomas originating from the parotid gland and from minor salivary glands.

Aberrant expression of membrane bound MUC1 and MUC4, which in normal tissues have an apical membrane localization, resulted in a distribution over the entire membrane highlighting the loss of membrane polarity in MECs[3]. In the parotid gland, MUC1 was expressed in low, intermediate and high grade MECs, while MUC4 was expressed in low and intermediate grade MEC but was absent in high grade MEC. In the minor glands, this difference in MUC1 and MUC4 expression was less pronounced.

The salivary mucin MUC5B and the gastric mucin MUC5AC were expressed in MEC of parotid and minor glands, particularly in low grades. The expression of the gel-forming secretory mucins, MUC5AC and MUC5B derived from the parotid gland is unique. The normal

parotid gland is a purely serous gland and therefore cannot express gel-forming secretory mucins. For the minor glands, this is less specific because minor salivary glands are often mixed (serous/mucous acini) and can express MUC5B but not MUC5AC. Intriguingly, MUC5B and MUC5AC expression was also found in intermediate and epidermoid type cells. These results are in line with previous studies with MECs and other malignancies[4,5].

The glycosylation machinery is frequently perturbed in cases of inflammation and in diseases such as cancer, resulting in the expression of truncated carbohydrate chains[6-9]. The expression of simple mucin-type Tn and T carbohydrates in low grade MECs was comparable to that in normal tissue, but decreased in high grade MECs. The sialyl-Tn antigen, however, was upregulated in MEC more frequently in the parotid gland than in the minor glands. Sialylation of Tn will terminate the further elongation of Tn to larger oligosaccharides, indicating that in MEC (mucin)glycoproteins are decorated with short, immature oligosaccharide sidechains. It has been suggested that the aberrant expression of mucin-associated antigens is a result of non-functional sugar-transferase such as T-synthase[10]. A specific molecular chaperone, *COSMC* , is required for the correct folding of T-synthase to form an active enzyme[11]. When *COSMC* is lost or mutated, the T-synthase is folded incorrectly and therefore becomes less active or even inactive, resulting in the expression of Tn antigens. The loss of *COSMC* function is only associated with the loss of T-synthase activity, and no obvious changes in other aspects of protein glycosylation have been observed[11].

The expression of mature carbohydrate epitopes Lewis^a and sulfo-Lewis^a was found in low grade MEC; Lewis^a was found in both mucous and in non-mucous cells and sulfo-Lewis^a primarily in mucous cells. Because of the observation of the expression of mature carbohydrate epitopes, Lewis^a and sulfo-Lewis^a, we concluded that the *COSMC* must be mutated and not deleted, otherwise no mature carbohydrate epitopes would be expressed.

Chromosomal copy number aberrations in mucoepidermoid carcinoma.

Chromosomal DNA changes such as the amplification and/or the deletion of a chromosomal region are frequent occurrences in cancer. Some tumor types can be characterized by chromosomal translocations, in which one specific region of a chromosome is switched with another region of a different chromosome. This is also true for a subset of MEC which have a t(11;19)(q21;p13) translocation. The presence of the translocation in MEC is associated with a more favorable prognosis[12,13].

In Chapter 3, a genome wide copy number aberration analysis was conducted by using micro-array comparative genomic hybridization (arrayCGH) technique on 27 MEC samples (10 translocation negative and 17 translocation positive).

Low/intermediate grade MECs had significantly fewer copy number aberrations than high grade MEC samples. Furthermore, translocation positive MEC samples had fewer copy number aberrations compared to translocation negative MEC samples. These results fit the

non-aggressive nature of low grade and translocation positive tumors, which generally have a more favorable prognosis than high grade and translocation negative MEC samples[12,13].

Within all 27 MEC samples, the most frequently gained regions were 16p11.2 and several regions on 8q , while the loss of region 1q23.3 (*RGS4*) was the most frequently detected loss. The most frequently detected copy number aberrations in low/intermediate grade MEC samples were the loss of 1p31.1, 1p31.1-p22.3, 12p13.2, and the gain of 16p11.2. The most frequently detected copy number aberration in high grade MEC samples was the loss of 8p23.3-p21.2. The most frequently observed copy number aberration in translocation negative MEC samples was the deletion of 3p14.1 (*FOXP1*), which was observed in 4 of the translocation-negative MEC samples. No recurrent copy-number-aberrations were found in translocation-positive MEC samples.

Based on the number of copy number alterations and translocation status, two subgroups could be made: (i) one group in which the majority of the samples were translocation-positive and had 6 or fewer copy number alterations, and (ii) a group in which most of the samples were translocation-negative and harbored multiple (>6) copy number alterations. This could suggest that there are different oncogenic pathways for MEC: one involving a relatively restricted aberration leading to the generation of a fused gene, and one involving more generalized chromosomal aberrations.

Differences in recurrent copy number aberrations have been reported in a number of other studies[14-16] and underscores the fact that heterogeneity in mucoepidermoid carcinoma is a common phenomenon. Only one region found in this study, the deletion of 9p21.3, was recorded by earlier studies[14-16]. The loss of 9p21.3, containing *CDNK2A/B*, is associated with an unfavorable prognosis and is also deleted in other salivary gland tumors such as adenoid cystic carcinoma and salivary duct carcinoma. Furthermore, the deletion of 9p21.3 is a frequent oncogenic event observed in head and neck squamous cell carcinomas and in lung cancer[17-19].

The recurrence of some chromosomal aberrations found in chapter 3, may suggest the involvement of certain specific genes. However, the genetic instability of these malignancies seems to be more important than a gain or loss of a specific chromosomal region(s) and/or gene(s).

Taken together, these results suggest that the number of copy number alterations in general, rather than aberrations in specific chromosomal regions may be used for differential diagnosis in MEC.

Salivary microRNA biomarkers for salivary gland tumors.

The altered gene expression due to cancer-associated changes in chromosomal DNA may impact the expression of RNA (mRNA and microRNA) and may be used for diagnostic purposes. Most research investigating the diagnostic value of mRNA and miRNAs is tissue-

based, but in recent years, salivary diagnostics has gained serious interest. It has evolved from the field of oral diseases[20,21] to the field of systemic diseases[22-24]. One of the new players in the salivary diagnostic field is miRNA. Salivary miRNA is both abundant and relatively stable. Taking into account the involvement of miRNAs in developmental processes, it is not surprising that several studies revealed tumor-specific miRNA expression patterns[26-29].

In chapter 4, the expression of salivary miRNA was explored to compare miRNA expression levels in whole saliva from patients with a parotid salivary gland tumor and miRNA expression levels in whole saliva from healthy individuals (Chapter 4). A combination of two miRNAs (hsa-miR-1233 and hsa-miR-211) was able to distinguish patients with a parotid gland tumor from healthy controls (sensitivity of 91%, and a specificity of 86%). Strikingly, the miRNA species that were detected in whole saliva were not present in the secretions from affected parotid glands. The differences in miRNA expression which we observed between unstimulated whole saliva (Table 3) and stimulated parotid saliva (Table 4) suggest that parotid saliva may have a totally different miRNA expression profile. Other studies have shown that tumors express miRNAs which may be involved in intercellular crosstalk. Tumor cells can transport genetic material, including miRNA, encapsulated in micro-vesicles (exosomes) to neighboring and/or distant cells, thereby affecting the miRNA expression of those cells and supporting cell growth and progression[30,31].

Additionally, differences in miRNA expression levels were determined in whole saliva from patients with a benign versus malignant parotid gland tumor (chapter 5). A combination of four miRNAs (hsa-miR-132, hsa-miR-15b, mmu-miR-140, and hsa-miR-223) could distinguish whole saliva samples from patients with a benign parotid gland and that from patients with a malignant parotid gland tumor (AUC 0.9; specificity of 95%; sensitivity of 69%).

Compared to other types of cancers, such as colon or breast cancer, salivary gland tumors are relatively rare, and only small sample sets are available. Because of the small sample size, the statistical power to make clear conclusions. There may be enough samples for a discovery phase but not enough to be validated in a larger independent sample set. In such cases, machine learning techniques may offer a suitable alternative to standard statistical methods. Although machine learning techniques are not new, they have only recently entered the field of cancer research. They can model high-dimensional and complex biological data which can have background noise and are often heterogeneous[32].

By application of machine learning techniques on the data set of chapter 5, an ensemble of seven miRNAs was identified (hsa-miR-374, hsa-miR-411, hsa-miR-599, hsa-miR-1285, hsa-miR-324-5p hsa-miR-449a and hsa-miR-449b) that distinguish between benign vs malignant salivary gland tumors (Chapter 6). These miRNA's are different form the ones emerging from standard statistical analysis, but had comparable diagnostic properties (AUC 0.9; specificity 86%; sensitivity 91%). Interestingly, four of these miRNAs can be associated with the regulation of the cell cycle, targeting genes that control the speed of cell division (Chapter 6, Figure 3). Hsa-miR-374 and hsa-miR-1285 (increased in saliva from patients with malignant

salivary gland tumor) speed-up the cell cycle, acting like gas pedals, while hsa-miR-449a and hsa-miR-449b (which are decreased in saliva from patients with benign tumors) decelerate the cell cycle, acting like brake pedals. One can envisage that increased levels of hsa-miR-374 and hsa-miR-1285 combined with decreased levels of hsa-miR-449a and hsa-miR-449b, may result in uncontrolled cell division.

In chapters 4 and 5, the research design was based on the prospective-specimen-collection, retrospective-blinded-evaluation or PRoBE design, which consists of discovery, verification and validation phases[33]. According to the PRoBE design, the discovered miRNAs are validated in an independent sample set[33]. Validating biomarkers in an independent sample set may prove difficult when the disease is very rare (such as salivary gland tumor) and/or the number of samples is limited. For these cases, the use of new statistical methods may be a good alternative for PRoBE. Unlike the PRoBE design, utilizing machine learning techniques require no new independent samples set is needed for the validation of the discovered miRNA. By splitting the sample set into a learning and a test sample set, the miRNAs can be statistically validated without needing new samples. Furthermore, unlike standard methods, statistical learning algorithms are able to capture complex group-based correlations in data which are difficult to detect using standard statistical testing[34-36].

Future perspectives

The studies compiled in this thesis were conducted with the goal of discovering additional biomarkers in order to increase the accuracy of diagnoses of salivary gland tumors. Our results suggest that saliva and miRNA expression profiling are good alternatives to the usually studied media (blood and tissue) and molecules (protein and mRNA) in cancer research.

An interesting prospect for future research would be the profiling the expression levels of miRNAs in whole saliva of the top 5 salivary gland tumor subtypes separately (pleomorphic adenoma, Whartins tumor, mucoepidermoid carcinoma, acinic cell carcinoma and adenoid cystic carcinoma). This group comprises about 80-90% of all salivary gland tumors. The miRNA data-sets can then be analyzed using machine learning techniques. This gives us predictive biomarkers as well as a more in-depth view of the pathology of salivary gland tumors.

Another interesting point is the fact that the expression level of diagnostic miRNA biomarkers decreases when the tumor is resected[37]. When a recurrence of the tumor occurs, does the expression level of the diagnostic miRNA biomarker increase again? Or are other miRNAs aberrantly expressed when the tumor is recurrent? This information could be beneficial for screening during the follow-up visits of patients with a salivary gland tumor.

However, before we can start profiling saliva from different salivary gland tumors, we need a bigger standardized salivary sample database. The Salivary Gland Tumor Biorepository (SGTB) in Houston, TX, USA (https://research.mdacc.tmc.edu/Salivary_DB/index.html) is

a good start. In this biorepository, fluids (saliva, NPBL, serum), tissue (FFPE and fresh) and slides of different SGT subtypes are stored. Researchers from all over the world can request samples for use in their research. The SGTB is an initiative of the National Institution of Health (USA), and is a good blueprint for a possible European SGT Biorepository.

While we're not quite at the level of Tri-corder readings, these results and techniques presented here can serve to expand both the scope and utility of molecular salivary diagnostics.

References

1. Seifert G, Brocheriou C, Cardesa A, et al. WHO international histological classification of tumours. Tentative histological classification of salivary gland tumours. Pathology Research and Practice. 1990; 186:555–81.
2. Schmidt RL, Hall BJ, Wilson AR, et al. A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions. American Journal of Clinical Pathology. 2011; 136:45-59.
3. Kufe D. Mucins in cancer: function, prognosis and therapy. Nature Review Cancer. 2009; 9:874-85.
4. Alos L, Lujan B, Castillo M, et al. Expression of membrane-bound mucins (MUC1 and MUC4) and secreted mucins (MUC2, MUC5AC, MUCH, MUC6 and MUC7) in mucoepidermoid carcinomas of salivary glands. American Journal of Surgical Pathology. 2005; 29:806-13.
5. Handra-Luca A, Lamas G, Bertrand JC, et al. MUC1, MUC2, MUC4, and MUC5AC expression in salivary gland mucoepidermoid carcinoma - Diagnostic and prognostic implications. American Journal of Surgical Pathology. 2005; 29:881-9.
6. Varki A, Cummings RD, Esko JD, et al. Essentials of Glycobiology Ch.9 (Cold Spring Harbour Laboratory Press, Cold Springs Harbour/New York, 2009).
7. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. EMBO Reports. 2006; 7:599-604.
8. Colpitts TL, Billing P, Granados E et al. Identification and immunohistochemical characterization of a mucin-like glycoprotein expressed in early stage breast carcinoma. Tumor Biology. 2002; 23:263-78.
9. Carneiro F, Santos L, David L, et al. T-(Thomsen-Friedenreich)-Antigen and Other Simple Mucin-Type Carbohydrate Antigens in Precursor Lesions of Gastric-Carcinoma. Histopathology. 1994; 24;105-13.
10. Ju TZ, Cummings RD. Identification of a unique molecular chaperone Cosmc-1 required for activity of the mammalian core 1 beta 3-galactosyltransferase. Proceedings of the National Academy of Sciences of the USA. 2002; 99:16613-8.
11. Aryal RP, Ju TZ, Cummings RD. The Endoplasmic Reticulum Chaperone Cosmc Directly Promotes in Vitro Folding of T-synthase. Journal of Biological Chemistry. 2010; 285:2456-62.
12. Behboudi A, Enlund F, Winnes M, Andrén Y, Nordkvist A, Leivo I, Flaberg E, Szekely L, Mäkitie A, Grenman R, Mark J, Stenman G. Molecular classification of mucoepidermoid carcinomas - prognostic significance of the MECT1-MAML2 fusion oncogene. Genes Chromosomes Cancer. 2006; 45:470-81.
13. Okabe M, Miyabe S, Nagatsuka H, et al. MECT1-MAML2 fusion transcript defines a favorable subset of mucoepidermoid carcinoma. Clinical Cancer Research. 2006; 12:3902-7.
14. Anzick SL, Chen WD, Park Y, et al. Unfavorable prognosis of CRTC1-MAML2 positive mucoepidermoid tumors with CDKN2A deletions. Genes Chromosomes Cancer. 2010; 49:59-69.
15. Jee KJ, Persson M, Heikinheimo K, et al. Genomic profiles and CRTC1-MAML2 fusion distinguish different subtypes of mucoepidermoid carcinoma. Modern Pathology. 2013; 26:213-22.

16. Zhang L, Mitani Y, Caulin C, et al. Detailed genome-wide SNP analysis of major salivary carcinomas localized subtype-specific chromosome site and oncogenes of potential clinical significance. American Journal of Pathology. 2013; 182:2048-57.
17. Rocco JW, Sidransky D. p16(MTS-1/CDKN2/INK4a) in cancer progression. Experimental Cell Research. 2001; 264:42-55.
18. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nature Review Cancer. 2006; 6:107-16.
19. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. English Journal of Medicine. 2008; 358:1118-28.
20. Park NJ, Zhou H, Elashoff D, et al. Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009; 15:5473–7.
21. Wong DT. Salivary diagnostics. Journal of the California Dental Association. 2006; 34:283–5.
22. Gao K, Zhou H, Zhang L, et al. Systemic disease-induced salivary biomarker profiles in mouse models of melanoma and non-small cell lung cancer. PLoS One. 2009; 4:e5875.
23. Parisi MR, Soldini L, Di Perri G, et al. Offer of rapid testing and alternative biological samples as practical tools to implement HIV screening programs. New Microbiology. 2009; 32:391–6.
24. Streckfus CF, Bigler LR, Zwick M. The use of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry to detect putative breast cancer markers in saliva: a feasibility study. Journal of Oral Pathology and Medicine. 2006; 35:292–300.
25. Zhang L, Farrell JJ, Zhou H, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology. 2010; 138:949–57.
26. Humeau M, Vignolle-Vidoni A, Sicard F, et al. Salivary MicroRNA in Pancreatic Cancer Patients. PLoS One. 2015; 10:e0130996.
27. Zahran F, Ghalwash D, Shaker O, Al-Johani K, Scully C. Salivary microRNAs in oral cancer. Oral Diseases. 2015; 21:739-47.
28. Xie Z, Yin X, Gong B, et al. Salivary microRNAs show potential as a noninvasive biomarker for detecting resectable pancreatic cancer. Cancer Prevention Research (Philadelphia, PA). 2015; 8:165-73.
29. Salazar C, Nagadia R, Pandit P, et al. A novel saliva-based microRNA biomarker panel to detect head and neck cancers. Cellular Oncology (Dordrecht). 2014; 37:331-8.
30. Kosaka N, Iguchi H, Yoshioka Y, et al. Secretory mechanisms and intercellular transfer of microRNAs in living cells. Journal of Biological Chemistry. 2010; 285:17442-52.
31. Valadi H, Ekström K, Bossios A, et al. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature Cell Biology. 2007; 9:654-9.
32. Koller D, Friedman N (2009) Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning. The MIT Press: Cambridge, Massachusetts, USA.
33. Pepe MS, Feng Z, Janes H, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. Journal of the National Cancer Institute. 2008; 100:1432-8.

34. Imangaliyev S, Keijser B, Crielaard W, et al. Online Semi-supervised Learning: Algorithm and Application in Metagenomics. IEEE BIBM. 2014
35. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference and prediction. 2nd edition. Springer: New York, New York, USA
36. Shalev-Shwartz S, Ben-David S (2014) Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press: New York, NY, USA.
37. Kodahl AR, Zeuthen P, Binder H, et al. Alterations in circulating miRNA levels following early-stage estrogen receptor-positive breast cancer resection in post-menopausal women. PLoS One. 2014; 9:e101950.

8.

Algemene Samenvatting

Speekselkliertumoren is een zeldzame groep van heterogene tumoren. Een diagnose van deze tumoren komt via verschillende diagnostische technieken (waaronder peroperatieve CT scan en histologisch onderzoek) tot stand. Het diagnosticeren van deze tumoren wordt bemoeilijkt doordat de verschillende subtypes overlappende histologische kenmerken hebben. Zoals met alle ziektes is het belangrijk dat de ziekte snel wordt gediagnosticeerd en behandeld. Hiervoor zijn (non-invasieve) biomarkers nodig. Biomarkers kunnen worden gevonden op verschillende moleculaire niveaus zoals DNA, RNA en eiwit. Het doel van deze studie was dan ook om moleculaire (non-invasieve) biomarkers te vinden die kunnen helpen in het diagnosticeren van speekselkliertumoren.

Naast de heterogeniteit is een ander nadeel, de zeldzaamheid van deze tumoren. Daarom is in eerste instantie gekozen om het onderzoek te verrichten in een van de meest voorkomende speekselkliertumoren, de mucoepidermoid carcinoma. Mucoepidermoid carcinomas zijn opgebouwd uit epidermoïde, intermediaire en mukeuze cellen. Deze laatste produceren een glycoproteïne mucine genaamd. Mucine heeft een grote diversiteit aan koolhydraat zijketens en kan worden onderverdeeld in membraangebonden en gel-vormende mucines. De expressie van sommige mucines en mucine-geassocieerde koolhydraat groepen is anders in tumoren dan in normaal weefsel (REFS). Voorbeelden hiervan zijn de herpositionering van membraangebonden mucines , *de novo* synthese van mucines en een slecht werkend glycosylering apparaat.

Deze observaties zien we ook terug in ons onderzoek naar in de expressie van mucines in mucoepidermoid carcinoma. Een interessante observatie in ons onderzoek was de expressie van gel-forming mucines, MUC5AC en MUC5B, in mucoepidermoid carcinomas uit de parotis klier. MUC5B is een speeksel mucine en omdat de parotis klier geen mukeuze acini bezit komt deze niet tot expressie in de parotis klier. MUC5AC is een mucine die onder normale omstandigheden niet tot expressie komt in geen een van de speekselklieren. De *de novo* expressie van MUC5AC is geobserveerd in vele subtypes pancreatobiliare neoplasma en wordt gezien als een early marker voor deze neoplasma (ref). De expressie van MUC5B en MUC5AC kan, net als bij pancreatobiliare neoplasma, een early marker zijn voor mucoepidermoid carcinomas die voortkomen uit de parotis.

Vele veranderingen die in een tumor tot stand komen, komen door veranderingen in het DNA zoals bijv. amplificatie of deletie van DNA, hyper-methylatie van promotoren, enz. dit kan leiden tot abnormale expressie van genen die tumorgroei stimuleren en de inhibitie van expressie van genen die tumorgroei tegengaan. Sommige diagnostische testen zijn gebaseerd op de amplificatie of deletie van specifieke regionen.

In totaal zijn er in ons onderzoek veel recurrent copy number aberraties gevonden. Echter deze werden in niet meer dan 5 van de 27 mucoepidermoid carcinoma samples gevonden. Veel laaggradige tumoren, veelal t(11;19)(q21;p13) translocatie positief, hadden weinig tot geen aberraties, terwijl meest hooggradige tumoren, veelal t(11;19)(q21;p13) translocatie negatief, meer dan 20 aberraties hadden. Het ziet er naar uit dat niet zozeer (een) specifieke

chromosomale regio(s) in mucoepidermoid carcinomas een diagnostisch waarde hebben als wel het aantal aanwezige aberraties.

Voor diagnostiek kunnen verschillende media (bijv. bloed, weefsel, urine en speeksel) worden gebruikt. Vooral speeksel diagnostiek is in de laatste jaren gegroeid. Het gebruik van speeksel als een diagnostisch medium heeft een aantal voordelen. Zo is verkrijgen van speeksel een non-invasieve handeling en bevat speeksel de meeste moleculen die ook aanwezig zijn in het bloed. Een ander voordeel is dat miRNAs rijkelijk en stabiel tot expressie komen in speeksel.

Daarom is er in de loop van het onderzoek een kleine verzameling van speeksels opgebouwd van patiënten met een speekselkliertumor. Door miRNA profielen in speeksels van patiënten met een speekselkliertumor te vergelijken met miRNA profielen in speeksel van individuen zonder een speekselkliertumor werden verschillen in miRNA expressie ondekt. Een combinatie van 2 miRNAs (hsa-miR-1233 and hsa-miR-211) kan een scheiding maken tussen speeksels van gezonde individuen en speeksels van patiënten met een speekselkliertumor onderscheiden (specificiteit 86%; sensitiviteit 91%). De ondekte miRNAs waren niet detecteerbaar in het parotisspeeksel van de patiënt met een parotisklier tumor. Wat kan duiden op een intercellulaire crosstalk. Intercellulaire crosstalk is een manier voor cellen om op afstand te communiceren, hetgeen kan resulteren in een verandering van de eiwitexpressie in de ontvangende cel. Na een verdere analyse van speeksel miRNA profielen van patiënten met een benigne of een maligne speekselkliertumor kan er op basis van 4 miRNAs (hsa-miR-132, hsa-miR-15b, mmu-miR-140 en hsa-miR-223) een onderscheidt tussen deze twee groepen worden gemaakt (specificiteit 95%; sensitiviteit 69%).

Een andere set van vier miRNAs (hsa-miR-132, hsa-miR-15b, mmu-miR-140, en hsa-miR-223) maakt het mogelijk om de tumor groep verder onder te verdelen in totaal speeksel van een patiënt met een maligne en een benigne speekselkliertumor.

De resultaten uit hoofdstuk 4, 5 en 6 zijn bemoedigend voor het ontwikkelen van een diagnostische speekseltest voor speekselkliertumoren. Verdere onderverdeling in maligne subtypes (mucoepidermoid carcinoma, acinuscelcarcinoom, adenoïd-cystisch carcinoom en adenocarcinoom) zou mooi zijn, maar aangezien speekselkliertumoren zeer zeldzaam zijn zou het moeilijk kunnen worden om het aantal speeksel samples bij elkaar te krijgen.

Een oplossing hiervoor kan liggen in de ontwikkeling van nieuwe statistische methode zoals machine learning. Bij machine learning analyseren algoritmes de data voor group-based connecties. In de afgelopen jaren is de hoeveelheid data dat word gegenereerd van enkele megabytes tot enkele gigabytes of terabytes gegroeid. Om data nog steeds hanteerbaar te houden, gebruiken de klassieke statistische technieken een filter (p-waarde, fold-change enz.). Hierdoor gaat een deel van de informatie die in de data zit verloren. Machine learning daarentegen plaatst geen filter, maar bouwt multivariate modellen waarbij alle data wordt gebruikt. Verder is machine learning een uitermate goede techniek om te gebruiken bij zeldzame ziektes (zoals speekselkliertumoren), omdat bij deze ziektes vaak sample sets klein zijn, terwijl er een grote hoeveelheid aan features (in dit geval miRNAs) wordt onderzocht.

Het onderzoek gepresenteerd in dit proefschrift wijst erop dat het diagnosticeren van speekselkliertumoren nog steeds een uitdaging is. Toch zouden speeksel miRNAs misschien wel de beste manier kunnen zijn om speekselkliertumoren de diagnosticeren. Het verkrijgen van speeksel is gemakkelijk en non-invasief. Bijkomend is het feit dat miRNAs stabiel zijn en niet snel zullen worden afgebroken in speeksel. Een andere interessant gegeven is het gebruik van machine learning in het onderzoek. Machine learning kan helpen om een hypothetisch inzicht te geven in de pathologie van speekselkliertumoren en andere zeldzame ziektes.

9.

Dankwoord

Hold on to what is good
even if it is a handful of earth,
Hold on to what you believe
even if it is a tree
which stands by itself,
Hold on to what you must do
even if it is a long way from here,
Hold on to life
even when it is easier letting go,
Hold on to my hand
even when I have gone away from you.

Taos Pueblo Prayer

Wetenschap is net als sport. Het ene staat meer in de spotlights dan het andere. Neem bijvoorbeeld het handboogschieten. Best wel een spannende en leuke sport om te zien, maar helaas alleen op tv een keer per vier jaar als de Olympische Spelen er weer zijn. Zo ook met speekselkliertumoren. Alleen op TV te zien als er een bekende voetbal coach (Tito Vilanova) er een laat verwijderen of Adam Yauch (Beastie Boys) en Tony Gwynn sr. eraan overlijden. Verder lijkt er weinig sexy aan het onderzoeken van speekselkliertumoren. En toch valt er nog zoveel te ontdekken en te onderzoeken! Tot deze conclusie kom ik na afloop van dit promotie traject. Zonder mijn promotoren Prof. dr. Bloemena en Prof. dr. Veerman en mijn copromotor Dr. Bolscher had ik dit nooit kunnen vermoeden. Dank hiervoor.

Collega's van Orale Biochemie, jullie waren geweldig. Er viel altijd wel wat te lachen bij de koffie of het thee momentje. Marjan, Kamran en Petra, jullie waren onmisbaar op het lab. Zonder jullie zouden de AIOs van Orale biochemie overgeleverd zijn aan de grillen van het onderzoek. Altijd goed gemutst in het lab, altijd tijd om wat na te rekenen of advies te geven. Stafleden Toon, Henk, Floris en Wim ondanks dat ik in mijn promotie onderzoek niet met jullie heb samengewerkt zag ik jullie toch elke dag bij de koffie. Jullie waren een gezellige groep collega met mooie verhalen van vroegere tijden.

En dan nog mijn collega Aio's van Orale Biochemie. Ja wat zal ik daar nu over zeggen of tegen zeggen. Ik heb nog altijd geweldige herinneringen aan onze kleine uitstapjes naar Leeuwarden, Utrecht en vooral naar de Jaap Eden IJsbaan met Aek en Parenaa. Tjitske, ja wij konden onze voorliefde voor fantasie boeken wel met elkaar delen en vooral als de verfilming niet strookte met wat er in de boeken stond. Bruin haar terwijl ze blond was etc. Ik wens je veel succes in alles wat er op je pad terecht komt. Irene, samen met Sabrina zijn jullie voor mij de belichaming van de song "Happy". Mocht het zijn voorgekomen dat jullie

ooit met een slecht humeur op het werk zijn geweest dan heb ik het dat nooit gemerkt. Altijd vol positiviteit. Soms zoveel om er jaloers op te zijn. Heel veel succes meiden. Nivedita, your determination and persistence still inspires me. I enjoyed cracking jokes with you and sometimes the occasional scientific sound-boarding. Wish you and your husband all the best and everything that's great in life! To the Thai students Sarkaw, Aek and Parenaa you were awesome. Enjoyed the time we shared the AIO room at ACTA. Aek, the way you looked at your first snow was and still is priceless. I hope that I can look at things never seen before with so much surprise and anticipation. I wish you all the best in Thailand. Pareena, you are just like the Dutch. Snow is fun and all but only for the first 3 min or so. Still we had a great time and wish you too all the best in life. Sarkaw enjoyed you having as a next door neighbor. May you all succeed in every field of life. Marjolein van Olderen je was een goede studente. Wat hebben we gelachen en gehuild samen. Gelukkig is alles goed op zijn pootjes gekomen en heb je nu 3 kidz. Wie had dat gedacht? Gelukkig draagt geen van hen grijs/paarse schoenen of mandarijnennetjes!

Nina, Greetje, Janice en Petra. Een bak koffie buiten en gaan met die banaan. Altijd lachen om (n)iets of iemand. Wat mij het meest is bijgebleven is de tunnel naar het hoofdgebouw waar iemand met krijt op de muur een woord had geschreven waar Nina niet meer van bijkwam. Meiden het gaat jullie goed. Weet zeker dat we zo nu en dan nog wel contact houden.

Natuurlijk zijn er nog altijd andere mensen om te bedanken voor hun adviezen en inzet. Zoals de mensen van de immunologie afdeling waar ik mijn glaasjes kleurden. Vooral Wim de Jong bedankt voor het snijden van mooie coupes. Verder wil ik via deze weg Dirk van Essen en dr. Ylstra van het CCA bedanken voor hun inzet in het CGH avontuur. Dirk bedankt voor de uitleg in het reilen en zeilen van arrayCGH experimenten. Dr. Ylstra bedankt voor uw inzet en adviezen bij het schrijven van het CGH manuscript. Prof. dr. Leemans wil ik zeer graag bedanken voor het mogen afnemen van speeksel bij patiënten met een speekselkliertumor. Zonder deze samples was hoofdstuk 4 in deze thesis niet mogelijk zijn geweest.

Evgeni and Sultan, I really enjoyed working with you guys. The meetings at TNO were intense, but there was always room for a little humor. You guys taught me machine learning, and I taught you guys some cancer and miRNA biology. I'll always have fond memories of you both, (especially when we tried to explain things to each other in terms of food, e.g. the pie analogies). I hope I get the chance to work with both of you again in the future. All the best to both of you.

In my time as a PhD student, I was honored to join Prof. David Wong's lab at University of California, Los Angeles during one of the coldest winters on record in the Netherlands. Dear Prof. Wong, I sincerely enjoyed the 3 months that I spent at your lab. Everyone in the lab made me feel right at home. Thank you for giving me the opportunity to take the lead in the salivary gland diagnostic project. My experience there helped me see how interesting salivary

research can be, and not only on a diagnostic level. It was a great learning opportunity for me. Furthermore, I would also like to thank dr. David Akin, dr. Janice Yoshizawa, dr. Michael Lee, dr. Nicolais Bonne and Samantha Wong. Janice, thank you for your guidance in the project; it was great working with you. I still have the red cardinal you made sitting here in the window sill. Nicolais, we had great times in Los Angeles. It was great to have a fellow European around. Let's meet up somewhere soon. Micheal and Samantha, it was great hanging out after work. Dinner and drinks were always a lot of fun and interesting. This work/learning trip was made possible with financial aid from the René Vogels Stichting (NVvO) and the Dutch Cancer Society (KWF).

I would also like to thank the entire Bailey/Anglada family for having me all these years. I felt right at home since day one. Grandpa Tito, Mrs Emily, Capt. T, Brian, Kimberly, Marsha, Eric, Jaiden en Sydney. Also the peeps that passed away. Grandma Lucy and Mrs Colleen. You are all awesome. Thanks for having me hang-out at Nana's Compound.

En dan zijn er nog een paar mensen over. Allereerst mijn paranimfen: Bertran Matse en Marjoleine Willems. Mar we kennen elkaar door en door en hebben super momenten meegemaakt en ook wat dalletjes. Je bent een van die mensen die zowat familie is geworden. Bertran, ja jij ben nu eenmaal familie. We kennen elkaar al 40 jaar jongen. Ik heb echt raak geschoten met een big bro als jij! Bedankt voor jullie hulp en support gedurende dit project.

Andere mensen die belangrijk zijn, zijn natuurlijk mijn ouders (Ton en Janny), Borus, Laura en Alan. Pa en ma bedankt voor jullie onuitputtelijke liefde en support. Ondanks dat het soms niet altijd meezat stonden jullie altijd voor de volle 100% klaar. Lau, jou lach en onuitputtelijke optimisme is een ware inspiratie. Veel liefde toegewenst samen met Bertran.

Alan. What can I say? No matter how many words I would write down on paper, no matter how dry and unpoetic, you would still edit them and somehow make them sing. Even these! Thank you for your loving and unselfish support throughout the years. It's finally time to build our home, no matter where that will be in the world.

Mocht ik mensen zijn vergeten mijn excuus daarvoor, maar tijdens je PhD ontmoet je zoveel mensen dat er altijd eentje vergeten wordt. Bedankt!